JMIR Public Health and Surveillance

Impact Factor (2023): 3.5
Volume 4 (2018), Issue 2 ISSN 2369-2960 Editor in Chief: Travis Sanchez, PhD, MPH

Contents

Original Papers

Reviews

Proposal

Corrigenda and Addendas

Letter to the Editor

Viewpoints

Original Paper

Incidence of Lower Respiratory Tract Infections and Atopic Conditions in Boys and Young Male Adults: Royal College of General Practitioners Research and Surveillance Centre Annual Report 2015-2016

Simon de Lusignan[1], MD[‡]; Ana Correa[1], MSc; Richard Pebody[2], PhD; Ivelina Yonova[1], MSc; Gillian Smith[3], MBBS; Rachel Byford[1], BSc; Sameera Rankiri Pathirannehelage[1], BSc; Christopher McGee[1,4], BSc; Alex J Elliot[3], PhD; Mariya Hriskova[1,4], BSc; Filipa IM Ferreira[1], PhD; Imran Rafi[5], PhD; Simon Jones[1], PhD

[1]Department of Clinical and Experimental Medicine, University of Surrey, Guildford, United Kingdom
[2]Respiratory Diseases Department, National Infection Service, Public Health England, London, United Kingdom
[3]Real-time Syndromic Surveillance Team, National Infection Service, Public Health England, Birmingham, United Kingdom
[4]Research and Surveillance Centre, Clinical Innovation and Research Centre, Royal College of General Practitioners, London, United Kingdom
[5]Clinical Innovation and Research Centre, Royal College of General Practitioners, London, United Kingdom
[‡]Simon de Lusignan

Corresponding Author:
Simon de Lusignan, MD
Department of Clinical and Experimental Medicine
University of Surrey
Leggett Building,
Guildford, GU2 7WG
United Kingdom
Phone: 44 1483 684802
Email: s.lusignan@surrey.ac.uk

Abstract

Background: The Royal College of General Practitioners Research and Surveillance Centre comprises more than 150 general practices, with a combined population of more than 1.5 million, contributing to UK and European public health surveillance and research.

Objective: The aim of this paper was to report gender differences in the presentation of infectious and respiratory conditions in children and young adults.

Methods: Disease incidence data were used to test the hypothesis that boys up to puberty present more with lower respiratory tract infection (LRTI) and asthma. Incidence rates were reported for infectious conditions in children and young adults by gender. We controlled for ethnicity, deprivation, and consultation rates. We report odds ratios (OR) with 95% CI, *P* values, and probability of presenting.

Results: Boys presented more with LRTI, largely due to acute bronchitis. The OR of males consulting was greater across the youngest 3 age bands (OR 1.59, 95% CI 1.35-1.87; OR 1.13, 95% CI 1.05-1.21; OR 1.20, 95% CI 1.09-1.32). Allergic rhinitis and asthma had a higher OR of presenting in boys aged 5 to 14 years (OR 1.52, 95% CI 1.37-1.68; OR 1.31, 95% CI 1.17-1.48). Upper respiratory tract infection (URTI) and urinary tract infection (UTI) had lower odds of presenting in boys, especially those older than 15 years. The probability of presenting showed different patterns for LRTI, URTI, and atopic conditions.

Conclusions: Boys younger than 15 years have greater odds of presenting with LRTI and atopic conditions, whereas girls may present more with URTI and UTI. These differences may provide insights into disease mechanisms and for health service planning.

(JMIR Public Health Surveill 2018;4(2):e49) doi:10.2196/publichealth.9307

KEYWORDS

population surveillance; medical record system, Computerized; general practice; pharyngitis; common cold; rhinitis; bronchitis; conjunctivitis; asthma; urinary tract infections; gastroenteritis

Introduction

The Royal College of General Practitioners (RCGP) Research and Surveillance Centre (RSC) publishes an annual report highlighting trends in respiratory, infectious, and some noninfectious conditions in England [1,2]. RCGP RSC data are extracted weekly from the computerized medical records (CMR) of >150 representative general practices in England, a sentinel network covering a population of over 1.5 million patients, that is, 3% of the population. RCGP RSC is one of the oldest sentinel networks; it was established as a Weekly Returns Service (WRS) in 1964, and it has just completed its 50th season of influenza surveillance [3,4]. The WRS reports continue to this day; they are available on the Web or individuals can sign up to receive them [2]. The network continues to conduct influenza surveillance and review vaccine effectiveness as part of a longstanding collaboration with Public Health England [5], which extends into broader public health interests, including promoting winter wellness [6]. RCGP RSC also contributes to European surveillance and studies of vaccine effectiveness [7,8].

RCGP RSC data are representative of the national population [1] in terms of the following:

1. *Age and gender of the population*: This is largely representative, although we have a slightly higher proportion of people aged 25 to 44 years and a lower proportion of people aged >75 years.
2. *Ethnicity*: The majority of patients in the RCGP RSC network were of white ethnicity (84.35%, 543,452/644,273). We have a slightly higher (though within 1%) prevalence of Asian and black ethnicities.
3. *Deprivation*: We have a slight over-representation of the less deprived using the Index of Multiple Deprivation (IMD). The mean deprivation score for the RCGP RSC population was 19.8 (SD 0.01), compared with the English population score of 21.8 (SD 0.0005).
4. *Other factors*: We have a geographical distribution of practices across England and actively recruit where we have gaps. A comparison of our practices with national pay-for-performance data, the quality and outcomes framework (QOF), and prescribing data suggests our data are representative. Our practices generally perform better at QOF (proportion of QOF targets achieved is 97.4% [529.58/559; SD 0.02%] compared with 94.7% [544.45/559; SD 0.0006%] for non-RCGP RSC practices). We also actively provide feedback to practices about recording of infections, especially encouraging accurate recording of episode type, to differentiate first or new (incident) cases from ongoing (prevalent) cases. We also extract data from all the different brands of the CMR system.

In addition to its weekly report, RCGP RSC produces an Annual Report (Multimedia Appendix 1) [9]. This report contains incidence data for 37 conditions or groups of conditions that are included in the WRS report. This report generally has a theme alongside these weekly incidence reports. In 2014/15, the theme was contrasting conditions that have different seasonal patterns [10]. In 2015/16, we explore gender differences in the presentation of respiratory conditions and infections in children and young adults. The network director hypothesized that in over 3 decades as a general practitioner (GP), he had seen more boys with lower respiratory conditions and asthma than girls, and this is the theme of the 2015/16 report.

This study explores differences, by age and gender, of respiratory conditions and infections presenting to general practice.

Methods

We extracted data from 155 participating practices who are members of RCGP RSC; a cohort of 1,589,702 patients registered during the period of 4 May 2015 to 8 May 2016. The data extracted were anonymized and encrypted; we only extracted coded data, not free text. Data were coded with Read version 2 or Clinical Terms version 3 [11].

RCGP RSC practices should have good data quality, particularly for influenza-like illness (ILI), acute infections, and respiratory conditions. RCGP RSC practices are encouraged to record the most likely diagnosis as a problem title and also assign an “episode type” to differentiate first or new presentations from ongoing care. Most of the data quality feedback to RSC member practices focuses on data quality for ILI, acute respiratory infection, and respiratory conditions. Since its inception, RCGP RSC has encouraged participating GPs to record valid and reliable diagnostic data; these approaches have been in place for some decades [12]. More recently, we have introduced financially incentivized training and practice-specific comparative feedback, which are modeled on the principles of audit-based education [13].

Data Processing and Analysis

We used the age bands that have been used long term by RCGP RSC to facilitate historic comparisons. These age bands were as follows: <1 year, 1-4 years, 5-14 years, and 15-24 years.

The WRS and Annual reports have traditionally reported 37 conditions or groups of conditions. On this basis, we excluded 5 because they were aggregates of several illnesses (but we kept lower respiratory tract infections [LRTIs] and upper respiratory tract infections [URTIs] as conditions of interest), 2 conditions that were not respiratory or infections, and 17 conditions that did not have sufficient sample size for our model (Figure 1). This left 13 conditions of the 37 items recorded, for which we explored gender differences in depth.

We grouped together LRTI and ILI, as influenza generally involves both upper and lower respiratory tracts, and diseases involving the lower respiratory tract are more clinically significant. We also grouped together URTI and conjunctivitis, as the latter is generally secondary to nasal obstruction. Finally, we grouped together asthma and allergic rhinitis (AR) as atopic respiratory conditions.

This left 13 conditions to consider in detail. Those were as follows: acute bronchitis and ILI (which we also grouped together with LRTI); acute tonsillitis, common cold, sinusitis, acute otitis media (AOM), and conjunctivitis (which we grouped together with URTI); asthma and AR, which we grouped together as atopic conditions; and finally, urinary tract infection

(UTI) and intestinal infectious disease. Although these 13 conditions have been given a single-disease label, we group together a number of codes that fit with that disease concept using an ontological approach [14]. For example, hay fever would be included within AR codes. A full list is provided in Multimedia Appendix 2.

To assess gender differences by age group, adjusting for other demographic variables, we developed 56 multivariate logistic regression models. The data were subset into 4 age bands used in RCGP RSC data (<1 year, 1-4 years, 5-14 years, and 15-24 years). Moreover, 14 conditions, including atopy (ie, AR or asthma), each measured per each age band, gave us 56 models. The outcome variable was whether the patient presented with each of the conditions above, and the explanatory variables included the following:

1. Gender (female was the reference group)
2. Ethnicity (white ethnicity was reference, and we divided into Asian (A), black (B), mixed (M), other (O), and unclassified (U) ethnicities) [15]
3. Deprivation was reported by the IMD quintile (quintile 1, the most deprived quintile, was used as reference)
4. Propensity to consult using consultation rate by decile band (band 1, the decile with the lowest consultation rate, was the reference, and band 6 was used in the models where there were no events in the consultation band 1; please refer to the information below for a description of this variable).

We also calculated the crude probability of males presenting with each condition in their respective age band and the adjusted probability based on the logistic model using the variables above.

Propensity to Consult

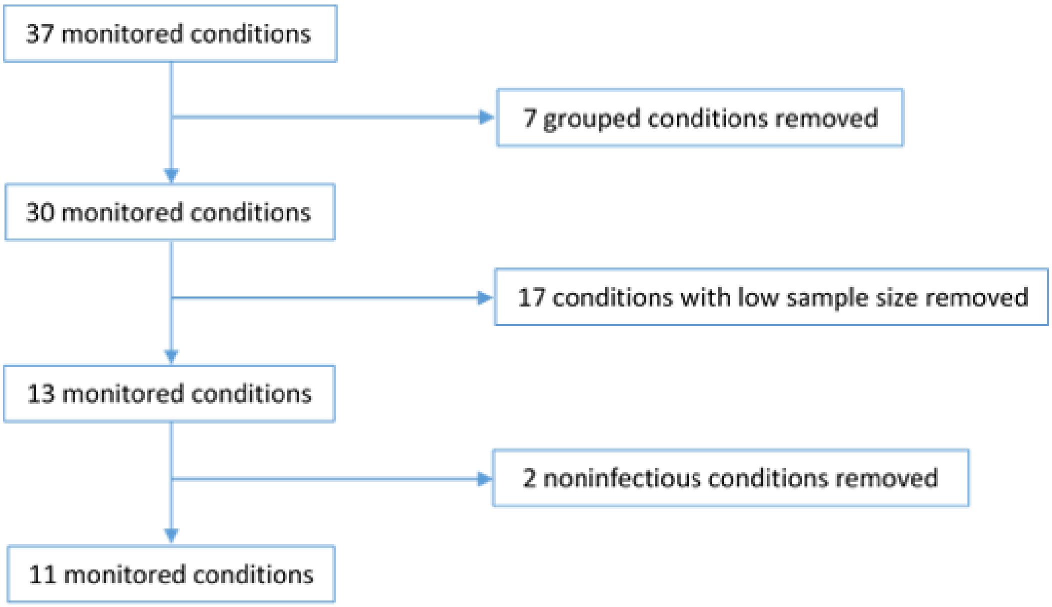
We assigned each patient a consultation rate. We then divided these rates into deciles, as a measure of propensity to consult. We calculated consultation rates in the following manner:

1. We summed up the number of visits (consultations or presentations) by patients to a GP as the number of consultations that they had in a year to find the numerator.
2. The denominator was 365, that is, the number of days in a year.
3. We divided the number of consultations per person in a year by 365 and multiplied the result by 100 to find the percentage of days in a year in which they consulted a GP.

The rate was defined as the number of different days that person attended their practice over 365 days. We conducted a sensitivity analysis, building 56 new regression models with an alternative measure for propensity to consult (annual rate of consultations without an “action” taken, in the form of a prescription or referral). This sensitivity analysis was conducted to determine whether removing consultations for patients who visited their GP without any action had an effect on the model.

The outcome variable and the other explanatory variables remained the same for these new regression models. The only changes were the different approaches to the “propensity to consult” explanatory variable. It must be noted that age, gender, and ethnicity may be confounding variables with consultation rates.

Figure 1. Selection of 13 conditions from the 37 items included in the Weekly Returns Service report and Annual Report of the Royal College of General Practitioners Research and Surveillance Centre for detailed investigation.



We report the results of the gender variable in this paper. For gender, we derived the odds ratio (OR), 95% CI, and probability (p) from the multivariate logistic regression [16]. An OR of >1 implies greater odds of a male/boy presenting with the condition; an OR of <1 suggests lower odds of a male/boy presenting, adjusting for other variables in the model. We created an aggregated table showing those conditions, with significant results denoted in italics (Table 1). Given the large number of models, we applied a Bonferroni-Šidák correction to the significance level of <.05, resulting in a new level of <.001 [17,18].

Additionally, the forest plots for each of the 13 conditions and the 4 age bands were reported separately (Multimedia Appendix 3). For each condition and each analysis, we quote OR, 95% CI, *P* value, and probability (p). Probability is calculated from the coefficients of logistic regression. We also include forest plots of the gender results of the 13 conditions closely studied, by age band. All the statistical analyses were conducted using functions in the statistical software R. The results of the sensitivity analysis were presented in a similar aggregated format (Table 2). We exclude a number of results in certain age subsets because of low numbers of events per predictor variables; we used a minimum of 20 events per variable as the threshold for exclusion [19].

The analysis presented in the Annual Report (Multimedia Appendix 1) includes the following:

1. Map of the national distribution of RCGP RSC practices
2. Summary tables showing the conditions we monitor, which is a new addition to the 2015/16 report:
   - Median age (using horizontal box-whisker plots)
   - Gender distribution of our monitored conditions
   - Ethnicity distribution comparing white and nonwhite
   - Median IMD (again using a horizontal box-whisker plot)
3. Week-by-week incidence of the conditions monitored by RCGP RSC—these data are published annually by RCGP RSC. Population denominators were based on the population registered in the participating practices for the study period. The weeks are numbered according to the International Organization for Standardization, ranging from week 1 to 52 in a single year [20].
4. An age-sex profile comparing the distribution of the disease with the national population, the distribution of the condition's deprivation score compared with the rest of RCGP RSC, and similarly for ethnicity. The level of deprivation was determined using IMD [21], scored from 0.5 (least deprived) to 92.6 (most deprived), based on each patient's Lower Super Output Area, which is determined from their postcode [22]. Ethnic groups, based on the 2011 English census categories, were assigned using an algorithm that incorporated proxy markers for ethnicity, such as language spoken [14]. This is also a new addition to the 2015/16 report.

Ethical and Data Sharing Considerations

Disease surveillance is part of standard health service activity, and therefore, no specific ethical approval was needed. The Health and Social Care Act 2012 includes the Secretary of State's duty to act to protect public health. RCGP RSC data extraction and analytics hub provides Public Health England with disease surveillance and vaccine effectiveness data. Data are pseudonymized as close to a source as possible, and no personal identifiers are held on the RCGP RSC secure network at the University of Surrey. We did not process the data of patients who had an "opt out" code.

Data are shared in a way which safeguards the confidentiality and anonymity of participants. Requests for access to data should be addressed to the data custodian of this study, Professor Simon de Lusignan.

Results

Lower Respiratory Tract Infections

Acute bronchitis was the single LRTI with the highest odds of presentation in boys up to the age of 15 years (Table 1). The greater ORs of boys attending were as follows: 1.59 (95% CI 1.35-1.87, *P*<.001) for those aged under 1 year, 1.13 (95% CI 1.05-1.21, *P*<.001) for those aged 1-4 years, and 1.20 (95% CI 1.09-1.32, *P*<.001) for those aged 5-14 years. After the age of 15 years, there are no significant gender differences (OR 1.11, 95% CI 1.00-1.23; *P*=.05).

Boys younger than 15 years had greater odds of presenting with LRTI and ILI, although much of this effect was due to acute bronchitis (Figure 2). For boys younger than 1 year, OR was 1.57 (95% CI 1.34-1.87; *P*<.001), for those aged 1-4 years, OR was 1.12 (95% CI, 1.05-1.20; *P*<.001), and for boys aged 5 to 14 years, OR was 1.20 (95% CI 1.10-1.31, *P*<.001). After the age of 15 years, no significant gender differences appear (OR 1.13, 95% CI 1.03-1.23; *P*=.01).

Upper Respiratory Tract Infections

Boys older than 5 years had lower odds of presenting with URTIs, with OR decreasing with age (Figure 3). For boys aged 5-14 years, OR was 0.89 (95% CI 0.86-0.92, *P*<.001), whereas for males aged 15-24 years, OR was 0.76 (95% CI 0.73-0.80, *P*<.001). Under the age of 5 years, there were no significant gender differences (<1 year: OR 1.04, 95% CI 0.96-1.13; *P*=.37 and 1-4 years: OR 0.99, 95% CI 0.95-1.02; *P*=.44).

Out of this grouping, tonsillitis had the lowest odds of presentation for boys aged 5-14 years (OR 0.81, 95% CI 0.76-0.85; *P*<.001) compared with common cold (OR 0.93, 95% CI 0.89-0.97; *P*<.001); the results for sinusitis showed no difference (Table 2). For males aged 15-24 years, sinusitis had the lowest odds of presentation (OR 0.60, 95% CI 0.52-0.70; *P*<.001), followed by tonsillitis (OR 0.73, 95% CI 0.69-0.77; *P*<.001) and common cold (OR 0.80, 95% CI 0.75-0.86; *P*<.001). The results for children aged 0-4 years were not significant, or, as in the case of sinusitis, the sample size was too small for analysis.

Atopic Respiratory Conditions

Compared with girls, boys aged 1-14 years had significantly greater odds of presenting with atopic respiratory conditions of AR and asthma (Figure 4). Under the age of 1 year, the sample was too small to conduct a regression (Table 3). For boys aged

1-4 years, OR of presenting with atopic respiratory conditions, compared with girls, was 1.37 (95% CI 1.16-1.61, *P*<.001); for boys aged 5-14 years, OR was 1.43 (95% CI 1.25-1.62, *P*<.001). For each individual condition, the gender differences were only significant for boys aged 5-14 years: OR 1.52 (95% CI 1.37-1.68, *P*<.001) for AR and OR 1.31 (95% CI 1.17-1.48, *P*<.001) for asthma.

The remaining conditions did not have significant results, with the exception of UTIs, where boys older than 1 year were less likely than girls to present with this infection (Table 4). Under the age of 1 year, the sample size was too low to run a regression. For boys aged 1-4 years, OR is 0.30 (95% CI 0.23-0.38, *P*<.001); for boys aged 5-14 years, OR is 0.14 (95% CI 0.11-0.18, *P*<.001); and for boys aged 15-24 years, OR is 0.04 (95% CI 0.03-0.06, *P*<.001).

Just under 1 in 20 boys <1 year presented with acute bronchitis (Table 1). The probability of a boy presenting with acute bronchitis decreased to about a third with each increasing age band (<1 year: 4.16%, 1-4 years: 1.42%, 5-14 years: 0.41%). For every 7 boys aged <1 year, 1 boy (15%) presented with a URTI (Table 2). There was a relatively little decrease until 5 years of age: 13.85%, 4.70%, and 2.45% for the 1-4 year, 5-14 year, and 15-24 year age bands, respectively. Although grouped with URTIs, conjunctivitis had a similar pattern of presentation to acute bronchitis, 4.48% in boys aged <1 year and 2.66% in boys aged 1-4 years. The probability of presenting with AOM peaked in the 1-4 year age band: incidence 2.51%.

Table 1. Odds ratios, 95% CI, *P* values, and probabilities of young males presenting with acute bronchitis, influenza-like illness (ILI), and lower respiratory tract infections and ILI (N=474,548).

Patients: age band and gender (N)	Acute bronchitis	ILI[a]	LRTI[b]+ILI
<1year male, ref=Female (N=14,066)			
Patients, n	723	9	730
OR	1.59	0.56	1.57
95% CI	1.21-2.08	0.05-6.03	1.20-2.06
P value	<.001[d]	.43	<.001[d]
Adjusted probability, %	4.16	0.16	4.35
1-4 years male, ref=Female (N=75,011)			
Patients, n	3794	187	3991
OR	1.13	1.04	1.13
95% CI	1.01-1.27	0.63-1.73	1.01-1.26
P value	<.001[d]	.79	<.001[d]
Adjusted probability, %	1.42	0.18	1.52
5-14 years male, ref=Female (N=211,752)			
Patients, n	1699	716	2426
OR	1.20	1.22	1.19
95% CI	1.02-1.40	0.86-1.72	1.03-1.38
P value	<.001[d]	.06	<.001[d]
Adjusted probability, %	0.41	0.31	0.47
15-24 years male, ref=Female (N=173,719)			
Patients, n	1835	380	2230
OR	1.11	1.18	1.14
95% CI	0.93-1.32	0.91-1.53	0.99-1.32
P value	.05	.04	.02
Adjusted probability, %	0.41	0.43	0.55

[a] ILI: influenza-like illness.

[b] LRTI: lower respiratory tract infection.

[c] OR: odds ratio.

[d] Indicates statistically significant, *P*<.001.

Figure 2. Forest plot of odds ratios of males younger than 25 years presenting with lower respiratory tract infection and influenza-like illness. Ages with statistically significant results are in italics. LRTI: lower respiratory tract infection.

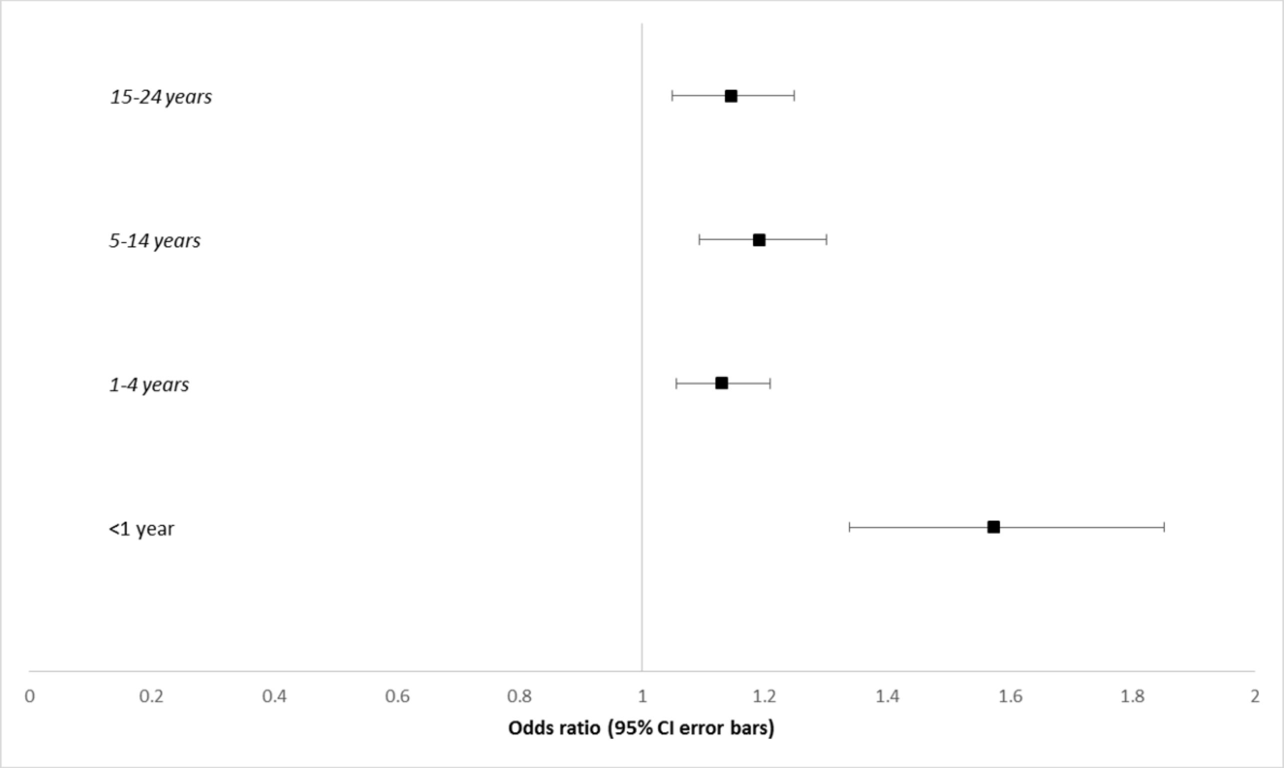


Figure 3. Forest plot of odds ratios of males younger than 25 years presenting with upper respiratory tract infection and conjunctivitis. Ages with statistically significant results are in italics. URTI: upper respiratory tract infection.

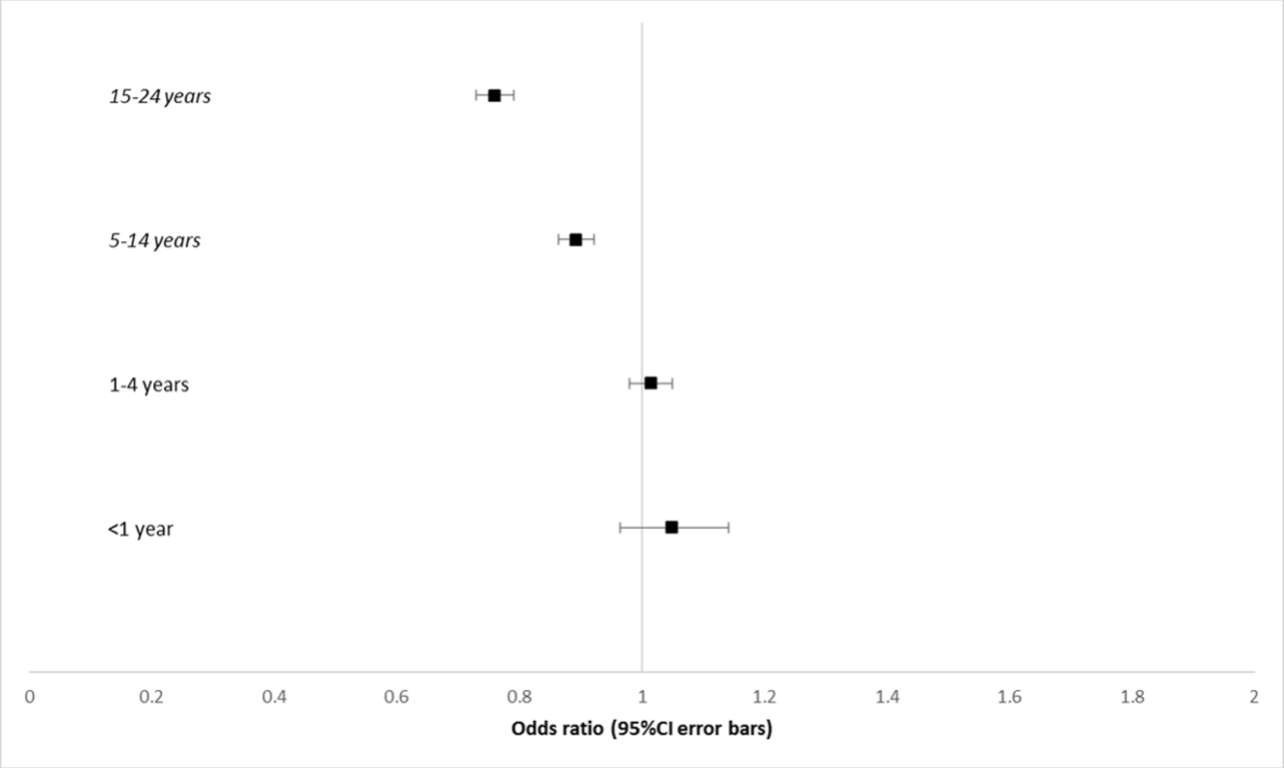


Table 2. Odds ratios, 95% CI, *P* values, and probabilities of young males presenting with upper respiratory tract infections and conjunctivitis (N=474,548).

Age band and gender (N)	Tonsillitis	Common cold	Sinusitis	Conjunctivitis	Acute otitis media	URTI[a]+conjunctivitis
<1year male, ref=Female (N=14,066)						
Patients, n	98	2607	1	1144	251	3527
OR[b]	1.22	1.09	N/A[c]	0.96	1.11	1.05
95% CI	0.61-2.45	0.94-1.27	N/A	0.78-1.19	0.71-1.72	0-91-1.21
P value	.34	.06	N/A	.55	.46	.26
Adjusted probability. %	0.05	11.00	N/A	4.48	0.78	14.81
1-4 years male, ref=Female (N=75,011)						
Patients, n	4144	15391	25	4592	5093	23654
OR	1.05	0.97	0.44	1.06	1.07	1.01
95% CI	0.95-1.18	0.91-1.04	0.11-1.76	0.96-1.18	0.97-1.18	0.96-1.07
P value	.11	.14	.05	.07	.03	.41
Adjusted probability, %	5.16	7.57	0.00	2.66	2.51	13.85
5-14 years male, ref=Female (N=211,752)						
Patients, n	5609	4598	942	937	931	12363
OR	0.81	0.93	0.92	1.05	0.93	0.89
95% CI	0.74-0.88	0.86-1.00	0.58-1.44	0.89-1.24	0.82-1.04	0.85-0.94
P-value	<.001[d]	<.001[d]	.54	.38	.03	<.001[d]
Adjusted probability, %	5.65	1.68	0.10	0.51	0.89	4.70
15-24 years male, ref=Female (N=173,719)						
Patients, n	6031	9036	221	1616	3468	2230
OR	0.73	0.80	0.60	0.96	0.83	0.76
95% CI	0.66-0.80	0.72-0.90	0.47-0.77	0.76-1.21	0.66-1.05	0.71-0.81
P value	<.001[d]	<.001[d]	<.001[d]	.58	.01	<.001[d]
Adjusted probability, %	2.98	0.77	0.20	0.15	0.29	2.45

[a] URTI: upper respiratory tract infection.

[b] OR: odds ratio.

[c] N/A: not applicable.

[d] Indicates statistically significant, *P*<.001.

The probability of a male child of 1 year or over presenting as an incident case of an atopic respiratory condition remains similar across age bands, just under 1 in 100 (0.80%-0.89%) each year (Table 3). Atopic incident cases are broadly equally split between AR and asthma, with the incidence of around 1 in 200 in these conditions individually. For the remaining conditions, the probabilities of boys presenting are under 1%.

Our sensitivity analysis produced very similar results to the original analysis (Multimedia Appendix 3). The only key differences were that OR of males aged 15 to 24 years attending with otitis media became significant (forest plots of the individual conditions are shown in Multimedia Appendix 4).

Figure 4. Forest plot of odds ratios of males younger than 25 years presenting with atopic respiratory conditions. Ages with statistically significant results are in italics. Age band under 1 has been excluded because of low sample size.

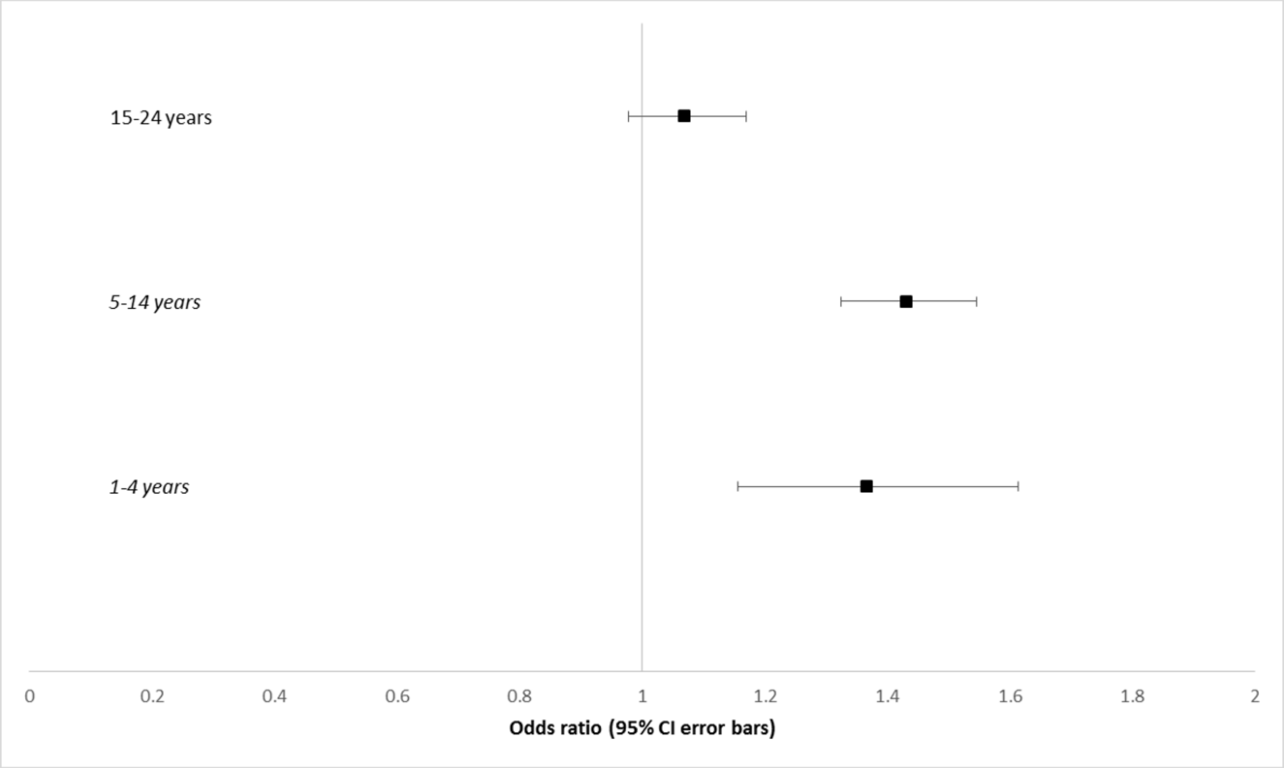


Table 3. Odds ratios, 95% CI, *P* value, and probabilities of young males presenting with atopic respiratory conditions of allergic rhinitis and asthma (N=474,548).

Age band and gender (N)	Allergic rhinitis	Asthma	Atopic respiratory disease
<1year male, ref=Female (N=14,066)			
Patients, n	4	1	5
OR[a]	N/A[b]	N/A	N/A
95% CI	N/A	N/A	N/A
P value	N/A	N/A	N/A
Adjusted probability, %	N/A	N/A	N/A
1-4 years male, ref=Female (N=75,011)			
Patients, n	277	368	636
OR	1.32	1.39	1.37
95% CI	0.87-1.99	0.97-2.00	1.04-1.80
P value	.03	.003	<.001[c]
Adjusted probability, %	0.30	0.40	0.80
5-14 years male, ref=Female (N=211,752)			
Patients, n	1361	946	2280
OR	1.52	1.31	1.43
95% CI	1.29-1.79	1.08-1.59	1.26-1.63
P value	<.001[c]	<.001[c]	<.001[c]
Adjusted probability, %	0.50	0.50	0.89
15-24 years male, ref=Female (N=173,719)			
Patients, n	1773	1243	2962
OR	1.14	1.00	1.07
95% CI	0.94-1.37	0.79-1.26	0.92-1.24
P value	.02	.95	.14
Adjusted probability, %	0.50	0.30	0.83

[a] OR: odds ratio.

[b] N/A: not applicable.

[c] Indicates statistically significant, *P*<.001.

Table 4. Odds ratios, 95% CI, *P* values, and probabilities of young males presenting with urinary tract infection and intestinal infectious disease (N=474,548).

Age band and gender (N)	UTI[a]	IID[b]
<1year male, ref=Female (N=14,066)		
Patients, n	35	152
OR[c]	0.71	0.88
95% CI	0.22-2.27	0.50-1.53
P value	.33	.45
Adjusted probability, %	0.28	0.67
1-4 years male, ref=Female (N=75,011)		
Patients, n	349	1088
OR	0.30	0.95
95% CI	0.19-0.45	0.77-1.17
P value	<.001[d]	.44
Adjusted probability, %	0.05	0.62
5-14 years male, ref=Female (N=211,752)		
Patients, n	1495	511
OR	0.14	1.07
95% CI	0.09-0.21	0.83-1.37
P value	<.001[d]	.41
Adjusted probability, %	0.04	0.14
15-24 years male, ref=Female (N=173,719)		
Patients, n	572	700
OR	0.04	1.05
95% CI	0.03-0.08	0.77-1.44
P value	<.001[d]	.60
Adjusted probability, %	0.03	0.10

[a] UTI: urinary tract infection.

[b] IID: intestinal infectious disease.

[c] OR: odds ratio.

[d] Indicates statistically significant, *P*<.001.

Discussion

Summary

Boys and young men were more likely to consult for LRTI and atopic conditions. Generally, the odds of presentation become less in older age bands, even if they remain significantly different. Our sensitivity analysis, excluding consultation where no action was taken, produced similar findings. In broad terms, the hypothesis that boys present more with LRTI and asthma up to puberty was supported by RCGP RSC data.

Strengths and Limitations

This paper used the data collected for RCGP RSC's WRS and Annual Report for the 2015-2016 season. RCGP RSC should have as good data quality as possible for infections, recognizing that routine data collected in the 10-min consultation has limitations [23]. Notwithstanding, the data quality of RCGP RS is considered the gold standard in primary care surveillance. This richness of research data available within a surveillance network can promptly produce initial insights into the epidemiology of a given disease, leading to more in-depth research.

Although the RCGP RSC network only covered 2.8% of the English population, it has been shown to be representative [1]. RCGP RSC is smaller than many of the other widely known UK primary care databases available for research. These include the Clinical Practice Research Data-link [24], the Health Improvement Network [25], and QResearch [26]. All have similarities, including the potential to link to other data such as hospital records and death data.

RCGP RSC differs in several ways. First, in their original form, most of the other databases were derived from a single brand of CMR system, whereas RCGP RSC extracts data from all the clinical systems. This creates issues around difference in version of coding system between brands [27] as well as related to their degree of problem orientation, which we adjusted for in our analysis [28]. Second, RCGP RSC data are probably the freshest of the data sources, perhaps inevitably so, given its surveillance function. Data extracted up to the end of the previous week are analyzed by Wednesday noon and are presented in the public domain by 2 PM on Thursday of the following week.

Comparison With Existing Literature

Although there is an overall propensity for males to consult less than females [29]; we did not see this in children or young adults. Generally, the literature around gender differences supports the hypothesis that boys are more prone to respiratory conditions and other infectious diseases, rather than this being a sociological phenomenon. There appears to be evidence that young males may be more susceptible to respiratory infections, respiratory symptoms, and hospitalization.

Acute respiratory infections, including tonsillitis, resulting in hospitalization of boys, have been shown to be 2.4 times higher in India [30]. An interesting study of wheeze from birth to adolescence showed a U-shaped difference in presenting with wheeze between boys and girls, with similar levels of presentation with wheeze of girls near birth and late adolescents. This is compatible with our findings that boys aged 1-14 years have an excess of asthma, a condition generally associated with wheeze [31]. Additionally, a questionnaire study reported that asthma and AR were more prevalent in boys aged 6 to 7 years, but with the exception of asthma, most health symptoms were more prevalent in girls aged 13 to 14 years [32]. A further study also described gender differences that were not explained by the presence of atopy [33]. A study of AR also reported a male excess in childhood, with females “growing into” AR in adolescence [34]. Our findings are also reinforced by a study of AR and asthma (in Brazil) in 6 and 7 year olds, reporting an excess of males [35].

The excess in infections in boys and young men may be reflected through into hospital admissions. Male vulnerability to infection, in terms of more boys admitted, has been seen across many disease categories (in Singapore) in which they report disparities of presentation to primary care [36]. A Danish study also showed disparity in admission in favor of boys, but that the reverse applied in adolescents and adults aged 15-25 years [37].

Not all studies support our findings. A study of URTIs including AR for ambulatory ear, nose, and throat practice found no gender difference in presentation [38]. Similarly, a study that followed up 294 children for a year showed no difference in rates of URTI or between genders [39].

The mechanism for such differences has only been hypothesized, with hormonal and epigenetic mechanisms having been proposed [40,41]. Differences in inflammatory markers might provide some insight. A study reported differences in inflammatory markers; white cell responses were longer in boys, whereas duration of fever was longer in girls [42].

Implications for Research and Practice

If it can be shown that boys and young adult males respond differently to infections than females and have different patterns of presentation; this will have important implications in health care provision across genders.

The preponderance of male presentations did not go away when we adjusted for the consultation rate (as a proxy of propensity to consult), nor when we adjusted for consultations at which no action was taken, as well as other key demographic variables. We see disease presentation as a complex bio-psychosocial phenomenon [43,44], and we conclude that although there may be important sociological contributions, it is more likely that genetic and hormonal reasons account for these differences in consultation rates.

Further research is needed at the specific disease level to explore this.

Conclusions

RCGP RSC has a long history as a surveillance network providing data about infections and respiratory illness. The Annual Report (Multimedia Appendix 1) contains important details about presentation across a wide range of conditions; it is extended in the 2015/16 report to include more details about the age-sex differences in disease presentation. Boys and young male adults appear to have greater odds of presenting with some infections and respiratory conditions. This phenomenon has not been reported from a substantial population group, such as RCGP RSC; confirmation of this observation and understanding its mechanism may enable us to tailor guidelines to gender differences.

Acknowledgments

The authors wish to thank the patients who consented to providing virology and other specimens in RCGP RSC practices, our member practices, and Apollo Medical Systems for managing secure data extraction. The authors would also like to thank Professor Mitch Blair for encouragement and support and Nick Andrews for helpful comments on the manuscript. Public Health England is the principal funder of RCGP RSC.

Authors' Contributions

All authors met the International Committee of Medical Editors criteria for authors and contributors. All authors were part of the scientific meetings to develop the RCGP Annual Report and this paper. SdeL was involved in the conceptual development of the paper and writing the main body with AC. AC carried out most of the statistical analyses. RP, GES, AJE, and IR were involved in review of analyses and editing of the paper. IY and MH were involved in practice liaison and editing of the paper. RB and SP were involved in data extraction and editing of the paper. CMcG was involved in statistical analysis and editing of the paper. FF was involved in research management and editing of the paper. SJ was involved in review of analyses, expert statistical advice, and editing of the paper.

Conflicts of Interest

SdeL is the principal investigator in a GlaxoSmithKline (GSK)-funded pilot to assess the potential to detect European Medicines Agency-listed possible brand-specific adverse events following immunization in near real time. He is the work package lead in Innovative Medicines Initiative (IMI)-funded project, Accelerated development of vaccine benefit-risk collaboration in Europe (ADVANCE) and a consortium member of Innovative Medicines Initiative (IMI)-funded project FLUCOP.

AC participated as a researcher in a GSK-funded pilot to assess the potential to detect European Medicines Agency-listed possible brand-specific adverse events following immunization in near real time.

The other authors have no conflicts of interest to declare.

Multimedia Appendix 1

This Annual Report is based on the data that we extract from more than 150 GP practices and draws together the principal elements of our work—disease surveillance, virological sampling, and vaccine effectiveness.

[PDF File (Adobe PDF File), 8MB - publichealth_v4i2e49_app1.pdf]

Multimedia Appendix 2

Read2 and CTV3 read code lists.

[PDF File (Adobe PDF File), 230KB - publichealth_v4i2e49_app2.pdf]

Multimedia Appendix 3

Odds ratios, 95% CI, P values, and probabilities of males presenting with infections and respiratory conditions by age band. This table is adjusted for whether any action was taken at the consultation. The * indicates statistically significant, *P*<.001.

[PDF File (Adobe PDF File), 28KB - publichealth_v4i2e49_app3.pdf]

Multimedia Appendix 4

Forest plots of the individual conditions; statistically significant results are in italics.

[PDF File (Adobe PDF File), 57KB - publichealth_v4i2e49_app4.pdf]

References

1. Correa A, Hinton W, McGovern A, van Vlymen J, Yonova I, Jones S, et al. Royal College of General Practitioners Research and Surveillance Centre (RCGP RSC) sentinel network: a cohort profile. BMJ Open 2016 Apr 20;6(4):e011092. [doi: 10.1136/bmjopen-2016-011092]
2. Royal College of General Practitioners. RCGP Research and Surveillance Centre 2018 URL: http://www.rcgp.org.uk/clinical-and-research/our-programmes/research-and-surveillance-centre [WebCite Cache ID 6yb2Xp0Rq]
3. Elliot AJ, Fleming DM. Surveillance of influenza-like illness in England and Wales during 1966-2006. Euro Surveill 2006;11(10):249-250. [Medline: 17130657]
4. de Lusignan S, Correa A, Smith GE, Yonova I, Pebody R, Ferreira F, et al. RCGP Research and Surveillance Centre: 50 years' surveillance of influenza, infections, and respiratory conditions. Br J Gen Pract 2017;67(663):440-441. [doi: 10.3399/bjgp17X692645] [Medline: 28963401]
5. Pebody R, Warburton F, Ellis J, Andrews N, Potts A, Cottrell S, et al. Effectiveness of seasonal influenza vaccine for adults and children in preventing laboratory-confirmed influenza in primary care in the United Kingdom: 2015/16 end-of-season results. Euro Surveill 2016 Sep 22;21(38). [doi: 10.2807/1560-7917.ES.2016.21.38.30348]
6. Gordon D, Bone A, Pebody R, de Lusignan S. The GP's role in promoting winter wellness. Br J Gen Pract 2017 Feb;67(655):53. [doi: 10.3399/bjgp17X688909] [Medline: 28126853]
7. de Lusignan S, Correa A, Ellis J, Pebody R. Influenza vaccination: in the UK and across Europe. Br J Gen Pract 2016 Sep;66(650):452-453 [FREE Full text] [doi: 10.3399/bjgp16X686677] [Medline: 27563114]
8. Broberg E, Melidou A, Prosenc K, Bragstad K, Hungnes O, WHO European Regionthe European Influenza Surveillance Network members of the reporting countries. Predominance of influenza A(H1N1)pdm09 virus genetic subclade 6B.1 and influenza B/Victoria lineage viruses at the start of the 2015/16 influenza season in Europe. Euro Surveill 2016 Mar;21(13) [FREE Full text] [doi: 10.2807/1560-7917.ES.2016.21.13.30184] [Medline: 27074657]

9. University of Surrey, Royal College of General Practitioners Research and Surveillance Centre. RCGP RSC Annual Report 2015/16 Supplementary Files URL: https://clininf.eu/index.php/rcgp-rsc-annual-report-2015-16-supplementary-files/ [WebCite Cache ID 6ue7MVF3T]
10. de Lusignan S, Correa A, Pathirannehelage S, Byford R, Yonova I, Elliot AJ, et al. RCGP Research and Surveillance Centre Annual Report 2014-2015: disparities in presentations to primary care. Br J Gen Pract 2017 Jan;67(654):e29-e40. [doi: 10.3399/bjgp16X688573] [Medline: 27993900]
11. de Lusignan S. Codes, classifications, terminologies and nomenclatures: definition, development and application in practice. Inform Prim Care 2005;13(1):65-70 [FREE Full text] [Medline: 15949178]
12. Fleming DM, Ayres JG. Diagnosis and patterns of incidence of influenza, influenza-like illness and the common cold in general practice. J R Coll Gen Pract 1988 Apr;38(309):159-162 [FREE Full text] [Medline: 3265157]
13. de Lusignan S. An educational intervention, involving feedback of routinely collected computer data, to improve cardiovascular disease management in UK primary care. Methods Inf Med 2007;46(1):57-62. [Medline: 17224982]
14. de Lusignan S. In this issue: ontologies a key concept in informatics and key for open definitions of cases, exposures, and outcome measures. J Innov Health Inform 2015;22(2). [doi: 10.14236/jhi.v22i2.170]
15. Tippu Z, Correa A, Liyanage H, Burleigh D, McGovern A, Van Vlymen J, et al. Ethnicity recording in primary care computerised medical record systems: an ontological approach. J Innov Health Inform 2017 Mar 14;23(4):920. [doi: 10.14236/jhi.v23i4.920]
16. Cox DR. Journal of the Royal Statistical Society. Series B (Methodological). In: The regression analysis of binary sequences (with discussion). Oxford: Wiley; 1958.
17. Bland JM, Altman DG. Multiple significance tests: the Bonferroni method. BMJ 1995 Jan 21;310(6973):170 [FREE Full text] [Medline: 7833759]
18. Abdi H. The Bonferonni and Šidák Corrections for Multiple Comparisons. Encyclopedia Meas Stat 2007;3:103-107 [FREE Full text]
19. van der Ploeg T, Austin PC, Steyerberg EW. Modern modelling techniques are data hungry: a simulation study for predicting dichotomous endpoints. BMC Med Res Methodol 2014 Dec 22;14:137 [FREE Full text] [doi: 10.1186/1471-2288-14-137] [Medline: 25532820]
20. International Organization for Standardization. Data elements and interchange formats -- Information interchange -- Representation of dates and times URL: https://www.iso.org/standard/40874.html [WebCite Cache ID 6yb3o68gx]
21. Department for Communities and Local Government. The English Indices of Deprivation 2015 URL: https://www.gov.uk/government/statistics/english-indices-of-deprivation-2015 [WebCite Cache ID 6yb3udsFT]
22. Office of National Statistics. Lower Layer Super Output Area (2011) to Ward (2015) Lookup in England and Wales URL: http://geoportal.statistics.gov.uk/datasets/lower-layer-super-output-area-2011-to-ward-2015-lookup-in-england-and-wales [WebCite Cache ID 6yb43RljC]
23. de Lusignan S, van Weel C. The use of routinely collected computer data for research in primary care: opportunities and challenges. Fam Pract 2006 Apr;23(2):253-263. [doi: 10.1093/fampra/cmi106] [Medline: 16368704]
24. Kousoulis AA, Rafi I, de Lusignan S. The CPRD and the RCGP: building on research success by enhancing benefits for patients and practices. Br J Gen Pract 2015 Feb;65(631):54-55 [FREE Full text] [doi: 10.3399/bjgp15X683353] [Medline: 25624277]
25. Blak BT, Thompson M, Dattani H, Bourke A. Generalisability of The Health Improvement Network (THIN) database: demographics, chronic disease prevalence and mortality rates. Inform Prim Care 2011;19(4):251-255 [FREE Full text] [Medline: 22828580]
26. Hippisley-Cox J, Stables D, Pringle M. QRESEARCH: a new general practice database for research. Inform Prim Care 2004;12(1):49-50. [Medline: 15140353]
27. Rollason W, Khunti K, de Lusignan S. Variation in the recording of diabetes diagnostic data in primary care computer systems: implications for the quality of care. Inform Prim Care 2009;17(2):113-119 [FREE Full text] [Medline: 19807953]
28. de Lusignan S, Liaw S, Dedman D, Khunti K, Sadek K, Jones S. An algorithm to improve diagnostic accuracy in diabetes in computerised problem orientated medical records (POMR) compared with an established algorithm developed in episode orientated records (EOMR). J Innov Health Inform 2015 Jun 5;22(2):255-264. [doi: 10.14236/jhi.v22i2.79]
29. Wang Y, Hunt K, Nazareth I, Freemantle N, Petersen I. Do men consult less than women? An analysis of routinely collected UK general practice data. BMJ Open 2013 Aug 19;3(8):e003320 [FREE Full text] [doi: 10.1136/bmjopen-2013-003320] [Medline: 23959757]
30. Krishnan A, Amarchand R, Gupta V, Lafond KE, Suliankatchi RA, Saha S, et al. Epidemiology of acute respiratory infections in children - preliminary results of a cohort in a rural north Indian community. BMC Infect Dis 2015 Oct 26;15:462 [FREE Full text] [doi: 10.1186/s12879-015-1188-1] [Medline: 26502931]
31. Tse SM, Coull BA, Sordillo JE, Datta S, Gold DR. Gender- and age-specific risk factors for wheeze from birth through adolescence. Pediatr Pulmonol 2015 Oct;50(10):955-962 [FREE Full text] [doi: 10.1002/ppul.23113] [Medline: 25348842]
32. Wieringa MH, Weyler JJ, Van Bever HP, Nelen VJ, Vermeire PA. Gender differences in respiratory, nasal and skin symptoms: 6-7 versus 13-14-year-old children. Acta Paediatr 1999 Feb;88(2):147-149. [Medline: 10102145]

33. Anderson HR, Pottier AC, Strachan DP. Asthma from birth to age 23: incidence and relation to prior and concurrent atopic disease. Thorax 1992 Jul;47(7):537-542 [FREE Full text] [Medline: 1412098]
34. Kurukulaaratchy RJ, Karmaus W, Raza A, Matthews S, Roberts G, Arshad SH. The influence of gender and atopy on the natural history of rhinitis in the first 18 years of life. Clin Exp Allergy 2011 Jun;41(6):851-859. [doi: 10.1111/j.1365-2222.2011.03765.x] [Medline: 21561494]
35. Castro LK, Cerci Neto A, Ferreira Filho OF. Prevalence of symptoms of asthma, rhinitis and atopic eczema among students between 6 and 7 years of age in the city of Londrina, Brazil. J Bras Pneumol 2010;36(3):286-292 [FREE Full text] [Medline: 20625664]
36. Hon KL, Nelson EA. Gender disparity in paediatric hospital admissions. Ann Acad Med Singapore 2006 Dec;35(12):882-888 [FREE Full text] [Medline: 17219000]
37. Jensen-Fangel S, Mohey R, Johnsen SP, Andersen PL, Sørensen HT, Ostergaard L. Gender differences in hospitalization rates for respiratory tract infections in Danish youth. Scand J Infect Dis 2004;36(1):31-36. [Medline: 15000556]
38. Dzięciołowska-Baran E, Gawlikowska-Sroka A, Mularczyk M. Diseases of the upper respiratory tract in preschool and school age children in ambulatory ear nose throat practice. Adv Exp Med Biol 2015;873:35-41. [doi: 10.1007/5584_2015_132] [Medline: 26318297]
39. Chonmaitree T, Revai K, Grady JJ, Clos A, Patel JA, Nair S, et al. Viral upper respiratory tract infection and otitis media complication in young children. Clin Infect Dis 2008 Mar 15;46(6):815-823 [FREE Full text] [doi: 10.1086/528685] [Medline: 18279042]
40. Del Principe D, Marconi M, Matarrese P, Villani A, Malorni W. Gender disparity in pediatric diseases. Curr Mol Med 2013 May;13(4):499-513. [Medline: 23438904]
41. Lazarus GM. Gender-specific medicine in pediatrics. J Gend Specif Med 2001;4(1):50-53. [Medline: 11324240]
42. Casimir GJ, Mulier S, Hanssens L, Zylberberg K, Duchateau J. Gender differences in inflammatory markers in children. Shock 2010 Mar;33(3):258-262. [doi: 10.1097/SHK.0b013e3181b2b36b] [Medline: 19543152]
43. Engel GL. The need for a new medical model: a challenge for biomedicine. Science 1997 Apr 8;196(4286):129-136. [Medline: 847460]
44. Balint M. The doctor, his patient, and the illness. Lancet 1955 Apr 2;268(6866):683-688. [Medline: 14354967]

Abbreviations

AR: allergic rhinitis
AOM: acute otitis media
CMR: computerized medical records
GP: general practitioner
IID: intestinal infectious disease
ILI: influenza-like illness
IMD: Index of Multiple Deprivation
LRTI: lower respiratory tract infection
OR: odds ratios
QOF: quality and outcomes framework
RCGP: Royal College of General Practitioners
RSC: Research and Surveillance Centre
URTI: upper respiratory tract infection
UTI: urinary tract infection
WRS: Weekly Returns Service

Edited by T Sanchez; submitted 01.11.17; peer-reviewed by S Mclean, X Sun; comments to author 04.01.18; revised version received 12.02.18; accepted 14.02.18; published 30.04.18.

Please cite as:
de Lusignan S, Correa A, Pebody R, Yonova I, Smith G, Byford R, Pathirannehelage SR, McGee C, Elliot AJ, Hriskova M, Ferreira FIM, Rafi I, Jones S
Incidence of Lower Respiratory Tract Infections and Atopic Conditions in Boys and Young Male Adults: Royal College of General Practitioners Research and Surveillance Centre Annual Report 2015-2016
JMIR Public Health Surveill 2018;4(2):e49
URL: http://publichealth.jmir.org/2018/2/e49/
doi:10.2196/publichealth.9307
PMID:29712621

Original Paper

A Surveillance Infrastructure for Malaria Analytics: Provisioning Data Access and Preservation of Interoperability

Mohammad Sadnan Al Manir[1], BSc, MS; Jon Haël Brenas[2], BSc, MS, PhD; Christopher JO Baker[1,3], PhD; Arash Shaban-Nejad[2], BSc, MPH, MSc, PhD

[1]Department of Computer Science, University of New Brunswick, Saint John, NB, Canada

[2]Oak Ridge National Laboratory Center for for Biomedical Informatics, Department of Pediatrics, The University of Tennessee Health Science Center, Memphis, TN, United States

[3]IPSNP Computing Inc, Saint John, NB, Canada

Corresponding Author:
Arash Shaban-Nejad, BSc, MPH, MSc, PhD
Oak Ridge National Laboratory Center for for Biomedical Informatics
Department of Pediatrics
The University of Tennessee Health Science Center
50 N Dunlap Street, R492
Memphis, TN, 38103
United States
Phone: 1 901 287 5836
Email: ashabann@uthsc.edu

Abstract

Background: According to the World Health Organization, malaria surveillance is weakest in countries and regions with the highest malaria burden. A core obstacle is that the data required to perform malaria surveillance are fragmented in multiple data silos distributed across geographic regions. Furthermore, consistent integrated malaria data sources are few, and a low degree of interoperability exists between them. As a result, it is difficult to identify disease trends and to plan for effective interventions.

Objective: We propose the Semantics, Interoperability, and Evolution for Malaria Analytics (SIEMA) platform for use in malaria surveillance based on semantic data federation. Using this approach, it is possible to access distributed data, extend and preserve interoperability between multiple dynamic distributed malaria sources, and facilitate detection of system changes that can interrupt mission-critical global surveillance activities.

Methods: We used Semantic Automated Discovery and Integration (SADI) Semantic Web Services to enable data access and improve interoperability, and the graphical user interface-enabled semantic query engine HYDRA to implement the target queries typical of malaria programs. We implemented a custom algorithm to detect changes to community-developed terminologies, data sources, and services that are core to SIEMA. This algorithm reports to a dashboard. Valet SADI is used to mitigate the impact of changes by rebuilding affected services.

Results: We developed a prototype surveillance and change management platform from a combination of third-party tools, community-developed terminologies, and custom algorithms. We illustrated a methodology and core infrastructure to facilitate interoperable access to distributed data sources using SADI Semantic Web services. This degree of access makes it possible to implement complex queries needed by our user community with minimal technical skill. We implemented a dashboard that reports on terminology changes that can render the services inactive, jeopardizing system interoperability. Using this information, end users can control and reactively rebuild services to preserve interoperability and minimize service downtime.

Conclusions: We introduce a framework suitable for use in malaria surveillance that supports the creation of flexible surveillance queries across distributed data resources. The platform provides interoperable access to target data sources, is domain agnostic, and with updates to core terminological resources is readily transferable to other surveillance activities. A dashboard enables users to review changes to the infrastructure and invoke system updates. The platform significantly extends the range of functionalities offered by malaria information systems, beyond the state-of-the-art.

(JMIR Public Health Surveill 2018;4(2):e10218) doi:10.2196/10218

KEYWORDS

malaria surveillance; global health; interoperability; change management; Web services; population health intelligence

Introduction

Malaria Surveillance Today

Malaria is an infectious disease with significant impact on developing countries. In 2016 alone, it caused 445,000 deaths worldwide, and globally around 216 million cases of malaria have been reported in 91 countries [1]. Populations in sub–Saharan African countries are most susceptible, with 80% of observed cases and recorded deaths worldwide [2,3]. Whereas a decrease in case incidence has been observed [1] since 2010, the rate of decline appears to have stalled, in part due to lack of adequate surveillance and intervention programs. An essential prerequisite for accelerating the decline of the disease and optimally targeting resources is an efficient surveillance infrastructure that can reliably deliver robust datasets. Poor data quality and sparseness and the coordination between different surveillance systems are ongoing challenges for the malaria surveillance community [4]. The increasing number of stakeholders, including international organizations, governments, nongovernmental organizations, and private sectors [5], that contribute to gathering the data can lead to siloed heterogeneous information systems and data sources that need to be integrated [6]. Overall, infrastructures for malaria surveillance are brittle [7], and stakeholders do not have sufficient confidence in the aggregated datasets to adopt any conclusions that could be meaningfully derived [8].

A comprehensive study [9] showed that malaria information systems that can collect, store, and analyze data and provide feedback to developers based on real-time information are few. Moreover, the most effective systems are still limited by the absence of real-time data aggregation, inconsistent decision support, and low levels of resolution, such as no mapping to households and no extrapolation across geographic borders. In specific cases where an information system [10-12] has been upgraded with improved visualization and reporting tools, other challenges remain unresolved. Data entered from field stations are managed centrally, and updates occurring at the central data source are not reflected in the field level immediately. Also, the options to generate reports are generally predetermined; hence, ad hoc queries cannot be answered without significantly overhauling the systems. Consequently, the types of surveillance queries that can be run to derive actionable knowledge in a timely manner are relatively few.

Furthermore, 11 widely used Web platforms were studied to assess how internet and Web technologies are used in the fight against malaria [13]. The elements of this study were focused on data, metadata, Web services, and categories of users. The results revealed that, although heterogeneous spatiotemporal malaria data came from multiple disciplines, they were rarely updated dynamically, no metadata were used to standardize them, Web services were inflexible for reuse and nonstandardized, and the platforms primarily served the scientific communities. The authors identified that, to improve these systems, interoperability through standardization is necessary. The World Health Organization’s global technical strategy [14] also identified that surveillance challenges in sub–Saharan African countries exist partly because mechanisms for facilitating data integration from distributed data silos and ensuring system interoperability are lacking.

Beyond the State-of-the-Art

Indeed, the malaria surveillance community is not alone in facing these challenges, and many other communities are investigating how to bring distributed datasets together in real time to support decision making. Researchers in other domains have sought to introduce guidelines for ensuring that source data are published in ways that ensure they are findable, accessible, interoperable, and reusable [15], albeit no specific technical solutions are proposed or mandated. In addition, standards and software prototypes have emerged specifically from the Semantic Web community [16] but, given the nature of these challenges, it is the combined benefits of applying guidelines and technical solutions that would provide a transformational development for surveillance practitioners.

The specific challenge of interoperability has two dimensions, namely structural and semantic interoperability. Structural or syntactic interoperability can be achieved by defining common syntax and formats for data exchange. For example, if two systems such as the Malaria Atlas Project [17] and Africa Development Indicators [18] are using demographic data from a census stored in another database, having interoperability between the two systems means that users of these databases can mutually access and reuse census data without having to store them locally or reformat the data on import. Semantic interoperability, on the other hand, is much harder to achieve. The goal is to ensure that the integrity and meaning of the data is preserved throughout the integration process. This can be achieved by mapping data to standardized vocabularies or terms in ontologies [19], such as the Vector Surveillance and Management Ontology (VSMO) [20], Infectious Disease Ontology-Malaria (IDOMAL) [21], or Mosquito Insecticide Resistance Ontology (MIRO) [22], which provide means to formally model a domain of interest using concepts in the domain, relations among the instances of concepts, and complex logical axioms. Two sets of data mapped to the same ontology terms or vocabularies can be regarded as having the same meanings. Ontologies facilitate semantic interoperability in integrated biomedical and health systems and can also be used to make malaria data sources comparable. Semantic interoperability achieved through mapping to community standard ontological terms is therefore an essential property for surveillance routines and enables more integrated access to data.

Overcoming the challenge of distributed data access has relied on established technologies such as Web services, but interoperability is still lacking in many implementations. In recent work, Web service-based data access and interoperability challenges have been tackled together using Semantic Web service infrastructures [23]. The benefits of this approach include enhanced findability of distributed online resources, easy construction of workflows to fetch data from multiple Web services, and access to source data in interoperable formats. Existing deployments of this approach [23,24] as part of the Semantics, Interoperability, and Evolution for Malaria Analytics (SIEMA) [23] platform have aimed to address the interoperability between multiple dynamic and evolving malaria data sources. Early trials have reported advanced query options [24] for end users without advanced technical skills. In addition, SIEMA seeks to manage infrastructure changes, where *change management* is defined as the process of preserving the integrity and consistency of an integrated system while keeping the information up-to-date. Here the primary activities during change management are detection, representation, validation, traceability, and rollback, as well as reproduction of the changes [25]. SIEMA has used authentic scenarios (use case) provided by the Uganda Ministry of Health to illustrate effectiveness in providing seamless access to distributed data and preservation of interoperability between online resources.

Objective

We introduce a prototype surveillance and change management platform, known as SIEMA, built from a combination of third-party tools, community-developed terminologies, and custom algorithms. We illustrate the methodology and core infrastructure used to facilitate interoperable access to distributed data sources using Semantic Automated Discovery and Integration (SADI) [26] Semantic Web services. We show a dashboard that reports on terminology changes that can render SADI services inactive, jeopardizing system interoperability, allowing end users to control and reactively rebuild services to preserve interoperability and minimize service downtime.

Methods

The SIEMA surveillance platform relies on the coordination and customization of a number of existing frameworks, and software and custom-developed algorithms. The architecture diagram in Figure 1 describes these resources in an abstract representation of the SIEMA platform first introduced by Brenas et al [23]. We briefly describe the key resources below.

Semantic Automated Discovery and Integration Semantic Web Service

SADI is a representational state transfer (RESTful) Web service framework that provides a set of conventions for creating Semantic Web services. The framework uses resource description framework schema (RDF[S]) [27] and Web Ontology Language (OWL) [28] standards for data representation and modeling, and HTTP-based recommendations (GET, POST) for interacting with the services. SADI services consume and produce RDF instances of OWL classes where the input data are decorated until they become an instance of the output OWL class. The services are deployed in a registry and can be automatically discovered, orchestrated, and invoked to return data in RDF on the query clients SHARE [29] and HYDRA [30,31].

Source Data and Standard Terminologies

Our research focuses on middleware for enabling discovery of datasets and tools for agile query composition, but deployments of these methodologies beyond prototypes will require full access to malaria data. Existing data repositories such as the Scalable Data Integration for Disease Surveillance [32] or Global Malaria Mapper [33] would be target resources for building service descriptions and including them in a generic registry. However, a recent study [13] suggested that in order to facilitate change management and semantic interoperability these data sources must be represented by standard terminologies related to their domain. Rather than defining new terminologies, it is a best practice to use community-adopted domain terminologies, if available, for defining the services. For this purpose, several ontologies are leveraged, including VSMO [20], IDOMAL [21], MIRO [22], and the Public Health Ontology [34]. The Clinical Data Interchange Standards Consortium [35] also offers ontologies.

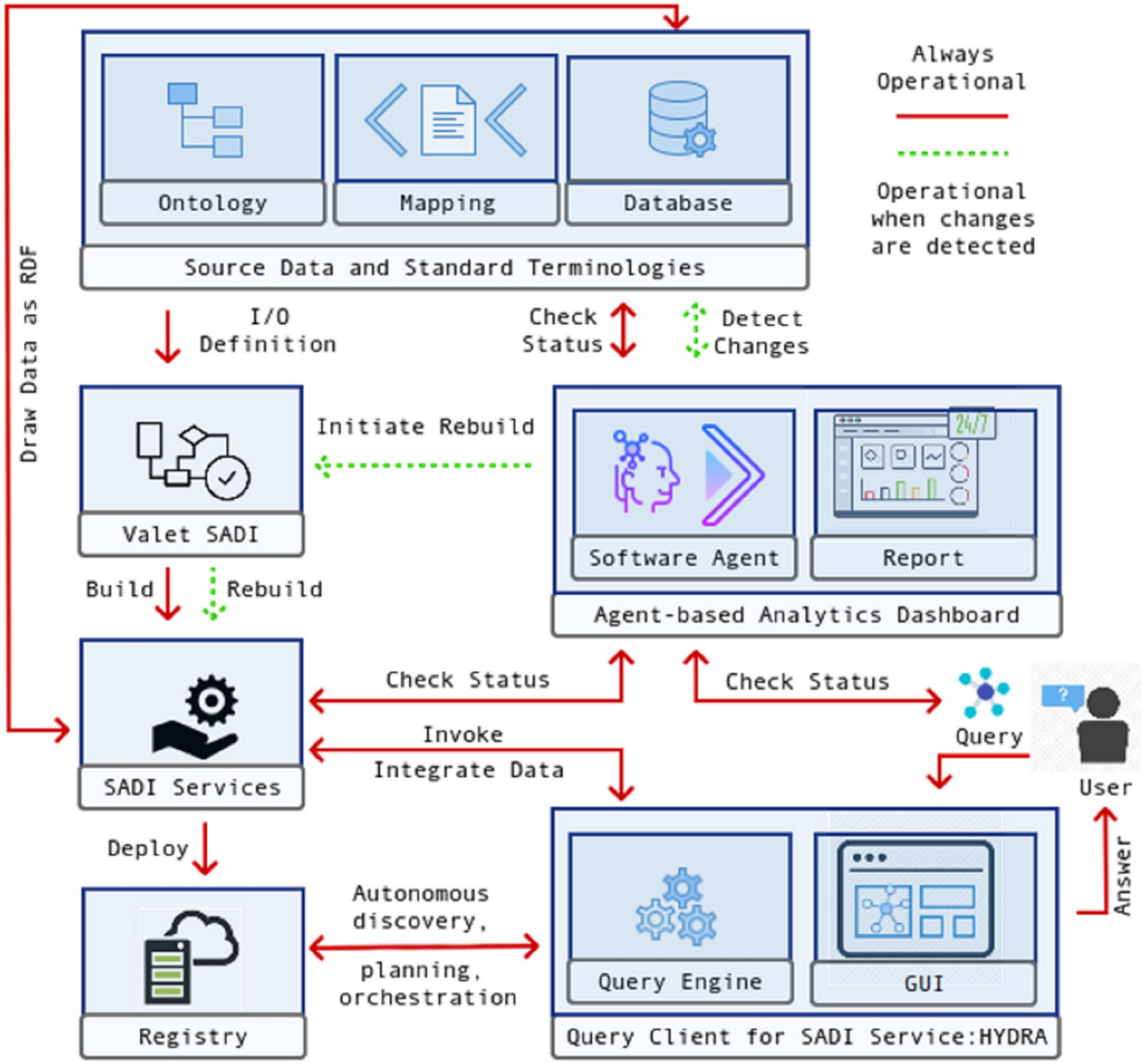
Query Clients for Semantic Automated Discovery and Integration Services

SHARE is a specialized open source query client that enables end users to discover, plan, and orchestrate SADI services in a registry and invoke them automatically from SPARQL [36] queries. The services are discovered by finding a match to the predicates attached to the input URIs after comparing both the semantic descriptions of the services' input and the output OWL classes. Each triple pattern in the SPARQL query is resolved by checking the predicates in the service registry. The order of invocation of services is strategized during the planning stage. The services are then orchestrated and invoked to execute the desired tasks. The output triples are produced through the binding of subject-object values to the corresponding variables in the SPARQL query. To generate the complete output, every triple in the query is resolved. HYDRA is a commercial query client for SADI services, developed by IPSNP Computing Inc. HYDRA can be used as a Java application programming interface, as a simple command line application, or through an intelligent graphical user interface (GUI) supporting ad hoc query composition by nontechnical users by combining Google-style keyword-based querying with query graph editing. Instead of queries being written in SPARQL, the GUI allows using a graph representation that is easier to understand.

Agent-Based Analytics Dashboard

Detecting changes in the source data schema, as well as the domain and service ontologies, is a prerequisite step for change management. Studies on the evolution of large domain ontologies [37] have shown that the vocabularies are subject to change [38] because they have shared authorship, and they evolve to represent new knowledge. Because of the high degree of dependency between system components in a relational data schema, even a small change in a data schema could result in cascading impacts across an application stack. In previous work, we presented features of a Web-based analytics dashboard for detecting and reporting changes [39].

Figure 1. Architecture of the Semantics, Interoperability, and Evolution for Malaria Analytics (SIEMA) surveillance framework. GUI: graphical user interface; I/O: input/output; RDF: resource description framework; SADI: Semantic Automated Discovery and Integration.



Underpinning the dashboard, software agents enable two key actions: (1) detecting changes and identifying their types, and (2) restoring the modified component to an operational state through repair and rebuilding. The types of changes we accommodate are addition (ie, extension), deletion (ie, obsoleting), and renaming (ie, refining) of components in domain ontologies and service ontologies [40,41], such as concepts, property restrictions, individuals, and axioms. Changes in the source data, such as tables and their attributes, data types, and indices, can also be reported [42]. Likewise, changes to saved queries and the corresponding service dependencies can be detected. Brenas et al [23] report further details of the agents designed to capture changes in various components of the infrastructure and their interaction with other software agents in the system.

Valet SADI

Valet SADI [43] generates SADI services automatically based on the formal definition of a service and rules for mapping a data source to ontology terms. The knowledge required to interpret the data is expressed using rules scripted in positional-slotted object-applicative (PSOA) RuleML [44]. These rules define how each element coming from a data source has to be matched to the vocabulary defined in the ontologies. Valet SADI, with the help of the domain ontologies and the semantic mapping rules, rewrites the declarative input and output definitions of a service into a query that is specific to the source data schema, such as Structured Query Language (SQL), and creates the complete Java code implementing the service functionality. It facilitates the generation and deployment of a service in a matter of seconds, which is essential in the context of change management, where uptime is a priority.

Results

Service-Based Querying of Malaria Data

The implementation presented in this paper is inspired by the objectives stated in the National Malaria Control Program by the Ugandan Ministry of Health [45]. The questions, shown in Textbox 1, based on these objectives resemble authentic queries for a surveillance practitioner in the malaria domain. In this research, we converted these questions into formal queries by semantically presenting them as one or more SADI services. This section illustrates how the queries can be represented using a combination of SADI services on a graphical query canvas of the HYDRA query engine.

Questions for Malaria Surveillance

We selected primarily a series of widely used malaria control interventions. The importance of indoor residual spraying has been well established by numerous studies throughout the world, especially in Africa [45,46], as an effective way of killing the mosquitoes that transmit malaria. Hence, the first question in Textbox 1 is about indoor residual spraying with a specific insecticide. We use this question as a running example in the rest of this paper. The use of insecticide-treated bed nets, alongside indoor residual spraying, is the second most effective intervention against malaria. Q2 in Textbox 1 is more complex to formalize and answer because it requires aggregating information about bed nets from geographic locations, insecticides used, and a comparison of the status of mosquito populations within a certain time frame in a specific location. Q3 in Textbox 1 anticipates that a malaria surveillance practitioner will be interested in predicting future trends and outbreaks to improve the allocation of resources where they are most needed. We consider these three queries as authentic target queries and adopted them as a basis for illustrating and evaluating the capabilities and appropriateness of the proposed technical approach.

Building Semantic Automated Discovery and Integration Services

To answers the questions, we created and deployed a list of SADI services in a registry. We focus below on how to create those services.

Source Data and Vocabulary

Vocabularies from one or more domain ontologies are used to define the input and the output of a service. The data schema of the source data is also necessary. The vocabularies and the data schema in Figure 2 are used to define the SADI services. In the domain ontology, the hierarchy of classes includes VSMO:0001957 and MIRO:10000239, each of which represents the concept of spraying an insecticide. The object properties *has_insecticide* and *located_in* link the records of spraying to the records of insecticides and geographic regions. The data property *has_name* represents the literal values of names of the specific spraying activity, the geographic region, and the insecticide used. In the data schema, the table labeled spraying contains information about all spraying activities; table geographicregion contains information about the locations where the insecticides were sprayed, and table insecticide contains insecticide information. In table spraying, the integer-valued attribute *id* is used to identify each spraying activity, attribute *name* represents the name of the activity, *location.id* refers to the location where the activity was performed, year represents the time it was performed, and *insecticide.id* refers to the insecticide used in the spray. In table geographicregion, the integer-valued attribute *id* represents each region and the attribute *name* represents the name of that region. Finally, in table insecticide, the integer-valued attribute *id* represents a pesticide, the attribute *name* identifies the name of the pesticide, and the attribute *mode.of.action* decides whether the pesticide is effective on mosquitoes that come in *contact* with it or both *contact & airborne*. The attributes *location.id* and *insecticide.id* act as foreign keys to the region and insecticide, respectively.

Description of Semantic Automated Discovery and Integration Services

The names of SADI services are expressed in two different forms: (1) allX, which retrieves all information regarding X without expecting any input, and (2) getYByZ, which retrieves Y based on the input Z. The input and the output of every service are defined in a service ontology using the terminologies from the domain ontologies. One such service is getInsecticideIdByIndoorResidualSprayingId, which takes an instance of spraying as input, which is any element whose *type* is indoor residual spraying. The service returns an *id* representing an insecticide. The input is decorated explicitly by the relation *has_insecticide*. Figure 3 defines the descriptions of this service.

Another service is getNameByInsecticideId, which takes an instance of an insecticide as input. The service returns a string as output, representing the name of the insecticide in the data, decorating the input by the relation *has_name*. Figure 3 shows the input-output descriptions of this service.

Textbox 1. Questions inspired by the National Malaria Control Program by the Ugandan Ministry of Health.

Q1. Which indoor residual sprayings used permethrin as an insecticide?

Q2. Which districts of Uganda that used permethrin-based long-lasting insecticide-treated nets in 2015 saw a decrease in *Anopheles gambiae* s.s. population but no decrease of new malaria cases between 2015 and 2016?

Q3. What are the future high-risk areas and at-risk time periods in Uganda?

Figure 2. Snapshot of source data schema and domain ontologies.

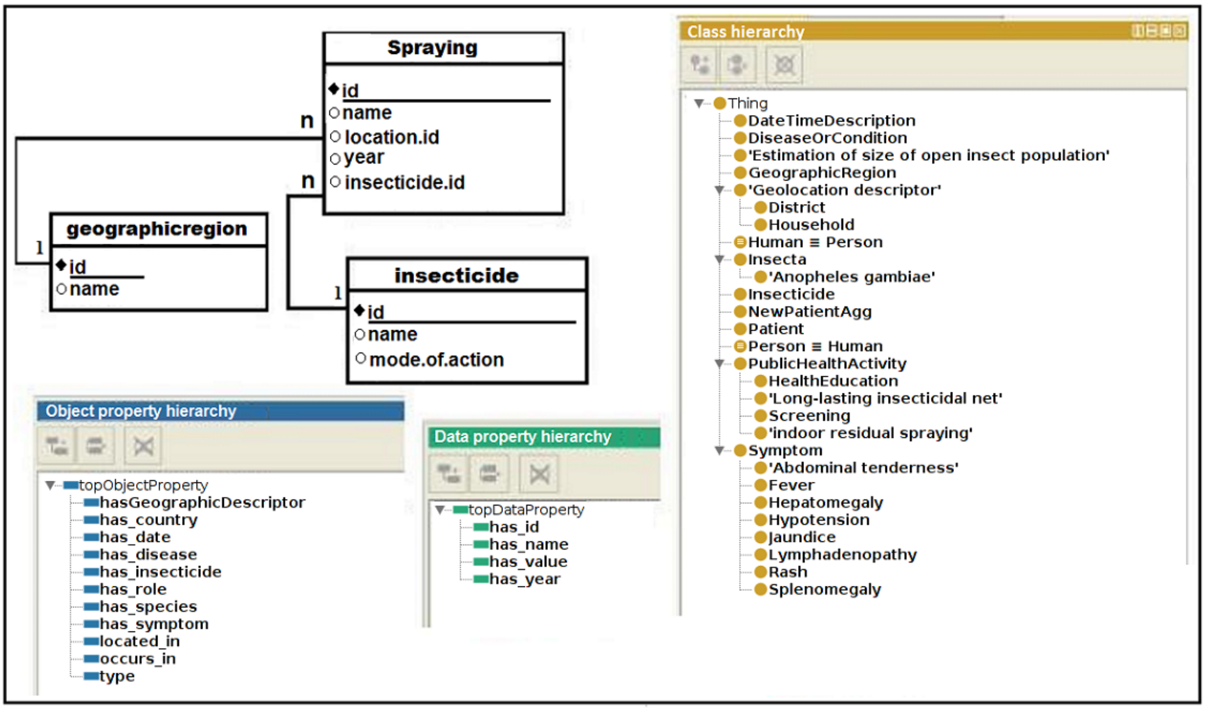
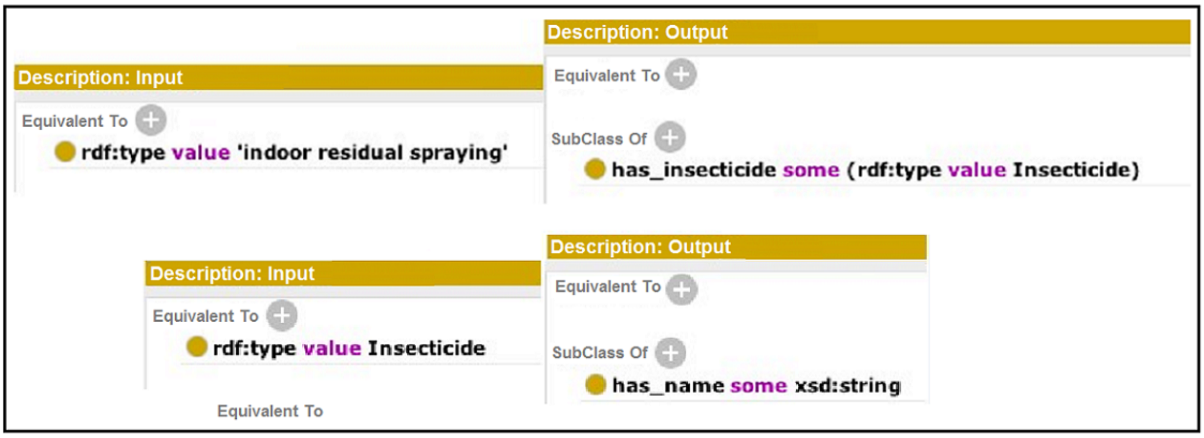


Figure 3. Input (left) and output (right) descriptions of the services getInsecticideIdByIndoorResidualSprayingId (top) and getNameByInsecticideId (bottom).



rdf:type value 'indoor residual spraying'

has_insecticide some (rdf:type value Insecticide)

rdf:type value Insecticide

has_name some xsd:string

Figure 4. A fragment of the registry of Semantic Automated Discovery and Integration (SADI) services.

Services

SADI Services

- getGeographicRegionIdByEstimationOfSizeOfOpenInsectPopulationId
- getDateByEstimationOfSizeOfOpenInsectPopulationId
- getValueByEstimationOfSizeOfOpenInsectPopulationId
- getSpeciesIdByEstimationOfSizeOfOpenInsectPopulationId
- getNameBySpeciesId
- getGeographicRegionIdByNewPatientAggId
- getDiseaseIdByNewPatientAggId
- getNameByDiseaseId
- getDateByNewPatientAggId
- getValueByNewPatientAggId

Textbox 2. Example of mapping rules in positional-slotted object-applicative RuleML.

```
Forall ?insecticideID (identityForInsecticideToinsecticideID(
            identityForInsecticide(?insecticideID)) = ?insecticideID)
Forall ?P (identityForInsecticide(identityForInsecticideToinsecticideID(?P))= ?P)
Forall ?id ?name ?mode.of.action (
 Insecticide(identityForInsecticide(?id)) :-
  db_insecticide(?id ?name))
Forall ?id ?name ?mode.of.action (
 has_name(identityForInsecticide(?id) ?name) :-
  db_insecticide(?id ?name))
```

Service Registry

Once the services have been generated, they are stored in a service registry. Figure 4 shows a fragment of the service registry that we use for malaria surveillance. HYDRA uses the registry to discover the services that will be called during the execution of the queries.

Provisioning of Semantic Automated Discovery and Integration Services

Specification and building of SADI services can be cumbersome, error prone, and tedious for nontechnical end users. Full implementation details are outside the scope of this paper; however, we encourage readers to look at the details in Brenas et al [23]. Here, we used Valet SADI, which accelerates the building of SADI services, to automatically implement the services from their input and output descriptions. Riazanov et al [30] demonstrated how rules expressed in PSOA RuleML semantically map the underlying source data schema to the vocabularies of the domain. As source data are usually not expressed in the language defined by the ontologies, such rules have been used to explain how to interpret the data. Textbox 2 shows an example of such mapping rules. Once these rules are available, Valet SADI uses them to access the data by rewriting the description of the services into correct queries for the mapped data schema.

Building Queries

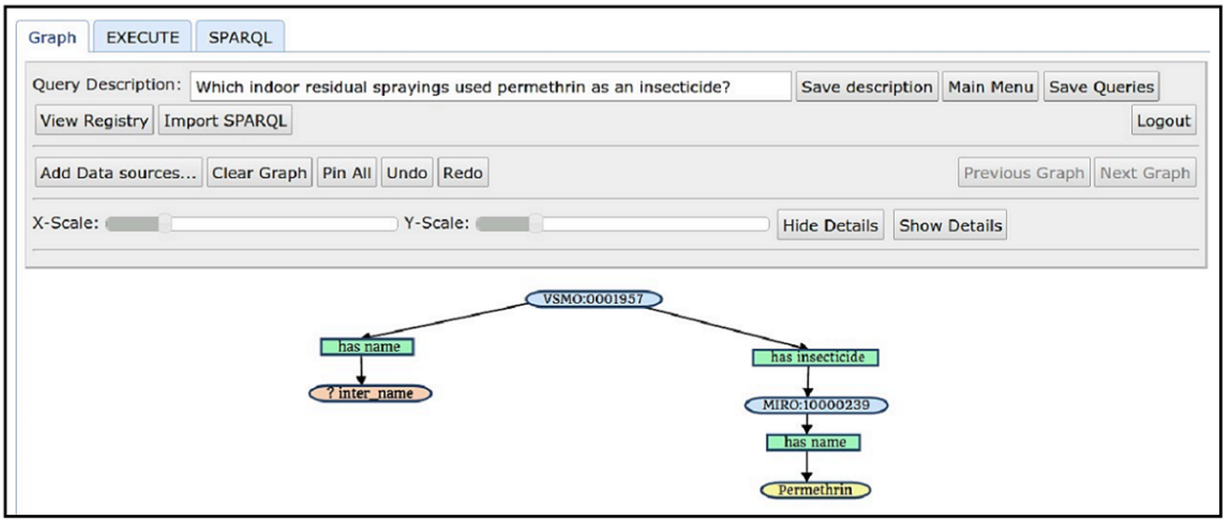
To illustrate query building, consider Q1 in Textbox 1, which is easily translated to a query. Figure 5 shows its graph representation. The GUI of HYDRA makes it easier to build queries because relationships are easier to understand and process in graph form than in SPARQL syntax.

Services in (3) and (4) are described above, while the service in (1) retrieves all identifiers of public health activities in Uganda, and the service in (2) retrieves the names of these activities. The branch on the right in Figure 5 uses the services (1), (3), and (4), while the left branch requires services (1) and (2). Queries are constructed in an incremental fashion. The root node starts with the service allPublicHealthActivities, which can then be extended either to the left or to the right. The service getNameByPublicHealthActivityId is used on the left to name the intervention represented by the variable *inter_name*. The root node is then decorated by the property *has_insecticide* with the description of the service getInsecticideIdByIndoorResidualSprayingId to represent an identifier of an insecticide. Finally, the service getNameByInsecticideId represents the name of the insecticide with a value *Permethrin*. As queries become more complex, the advantage of graph form over the raw SPARQL syntax becomes evident. Figure 6 shows the graph form of Q2. Although the query graph is significantly more complex and larger than the query in Q1, it is much easier to understand for any user than the raw SPARQL representation. Due to space constraints, we do not describe each service used in the query graph of Q2 or of Q3.

The query in Figure 5 calls on four distinct SADI services:

1. allPublicHealthActivities,
2. getNameByPublicHealthActivityId,
3. getInsecticideIdByIndoorResidualSprayingId, and
4. getNameByInsecticideId.

Figure 5. Graph representation of a query for the question “Which indoor residual sprayings used permethrin as an insecticide?" prepared on the HYDRA graphical user interface.



Query Description: Which indoor residual sprayings used permethrin as an insecticide?

Figure 6. Graph representation of a query for the question “Which districts of Uganda that used permethrin-based long-lasting insecticide-treated nets in 2015 saw a decrease in *Anopheles gambiae* s.s. population but no decrease of new malaria cases between 2015 and 2016?”.

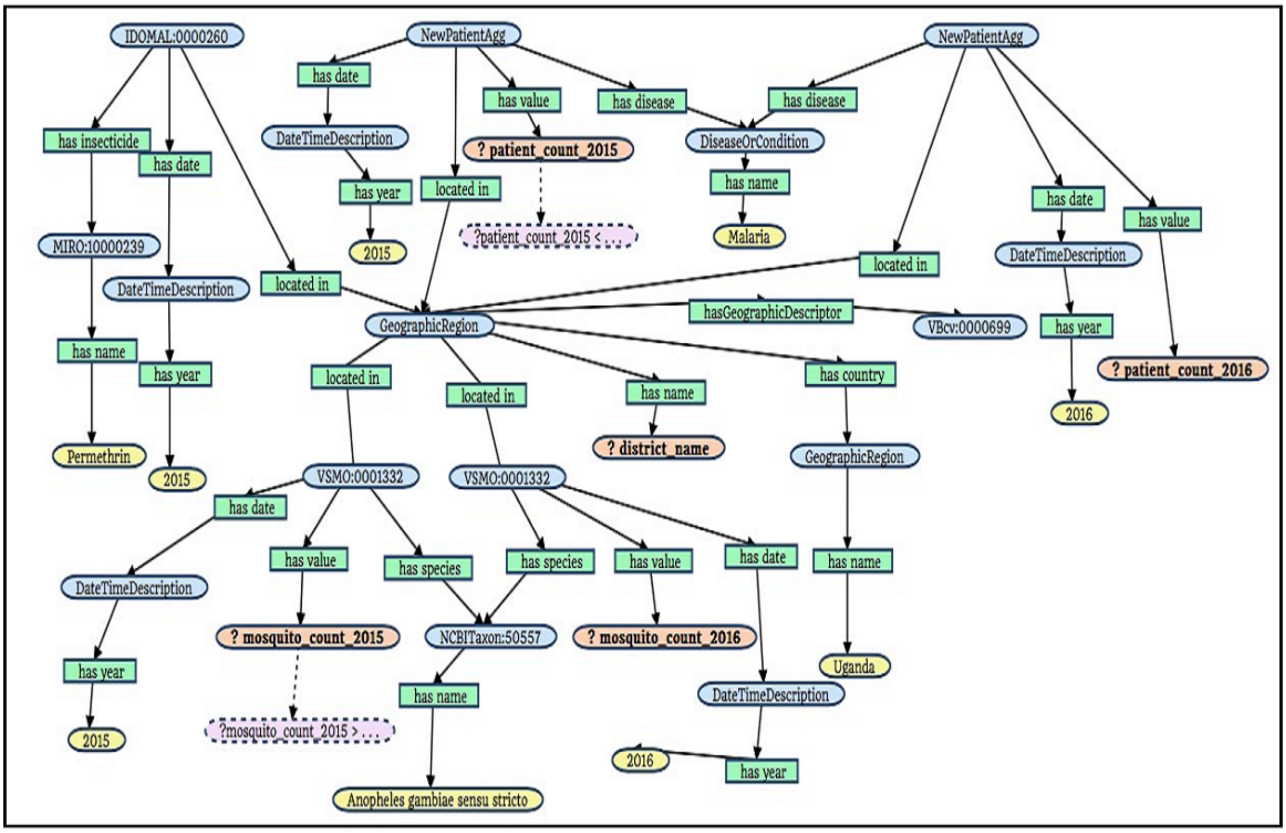
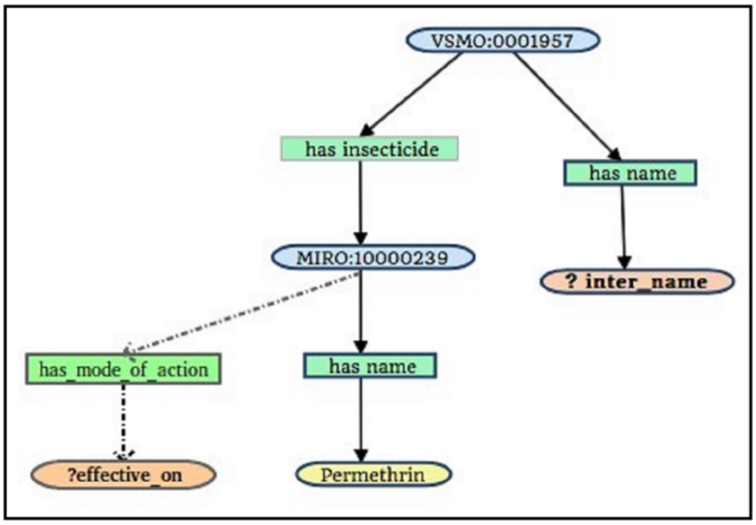


Figure 7. Graph representation of the query “Which indoor residual spraying used permethrin as an insecticide and which kind of mosquitoes will be affected by it?”.



Change Management

The previous section outlined how, using Semantic Web services, it is possible to answer complex questions relevant to malaria surveillance. Special attention is needed before considering the introduction of a new methodology in a dynamic context where data and middleware are not static. Several possible changes that could occur in a malaria surveillance framework have been described and classified according to the degree to which they affect data access and their likelihood to affect interoperability of the system [23]. In the following, we illustrate an approach to address the notion of change management through the introduction of change detection tools and triggers for the reactive rebuilding of service to ensure service uptime. Specifically, we discuss the addition of terms that are used in service ontologies and the use of a reporting tool for detecting and displaying these types of changes.

Changes in the Service Ontologies

Whenever a definition of an existing service is modified, the associated service ontology is changed, but the code implementing the service remains unchanged. As a result, when a service is invoked during the execution of a query, it does not return the anticipated output because the terms used in the code are incompatible with the new definition in the service ontology. In the SIEMA framework, the change capture agent implemented within the dashboard detects the changes in the terms used in the service ontology by comparing the modified version of the ontology with the one it was modified from. The role of the change capture agent can be illustrated in the case of a term addition.

To illustrate a scenario involving the addition of new terms, consider the query in Figure 5 again. The mode of action, mentioned in the World Health Organization’s recommendation [47], plays an important role in choosing insecticides for indoor residual spraying against malaria vectors.

Let us assume that the end user is interested in the query “Which indoor residual spraying used permethrin as an insecticide and which kind of mosquitoes will be affected by it?” Figure 7 shows the graph representation of such a query. To be able to answer this query, the existing service getNameByInsecticideId can be modified by adding the term *has_mode_of_action* and the datatype xsd:string that it returns. The new output description in Figure 8 is defined by adding the two terms to the old description.

A change capture agent detects the changes of two terms, identifies them as addition, and displays them on the dashboard in a tabular form. Table 1 shows the consolidated information about the changes on the dashboard, which displays the time the changes were detected in the *Timestamp* column; the textual description of the change in the *Description of change* column; the types of changes identified as either addition in the *Entity added* column, deletion in the *Entity deleted* column, or new name in the *Entity renamed* column; the list of services affected in the *Affected service* column; and the list of queries affected in the *Affected query* column.

Status of Services

At any time, the status of all deployed services in the registry is displayed on the dashboard. Services can be either active, which can be used in queries, or inactive, which need to be repaired before using them again. Figure 9 displays a list of active and inactive services in green and red, respectively. The URI of the service is tabulated in the *Service URI* column, the description is under *Description*, the time the service was created is under *Time of creation*, and the time the service was rebuilt is under *Time of rebuild*. The final column is *Request rebuild*, which allows for placing an inactive service in a queue to be rebuilt and subsequently redeployed in the service registry. The getNameByInsecticideId service is shown in red because it became inactive due to the addition of terms in its output definition.

Reacting to Changes

The addition of terms to the definition of an active service renders the service description incompatible with the target functionality and existing service code, and renders associated queries dysfunctional. To resolve the inconsistency, it is necessary to repair and rebuild the services in line with the new requirements. Specifically, the end user now wants to access the data that were not previously available from a service, namely, in this example, the mode of action of an insecticide. It is thus necessary to reimplement the service corresponding to the altered service description ensuring that the domain ontologies, the data schemata, and the PSOA semantic mapping rules underpinning the service are accurate and will support the new target functionality.

Given that a data resource contains the information about the mode of action of insecticides, the key question is whether the semantic mapping rules already map those data to an existing concept or relation of the domain ontologies. If that is the case, then all components required to rebuild the services exist, and it is possible to proceed to the next step in the Valet SADI rebuild. Otherwise, it is necessary to identify missing rules and add them, or extend a local domain ontology with a missing concept or relation that exists in the service ontology. Once this is done, a rule must be created to define a new mapping and to make rebuilding the service possible.

Rebuilding the Services Using Valet SADI

By leveraging Valet SADI’s autogeneration capability, the damaged service can be quickly rebuilt and deployed once changes are detected and identified, and a rebuild is requested. To illustrate this, we refer to the query shown in Figure 7, which calls services retrieving data from the source database schema and list the changes below.

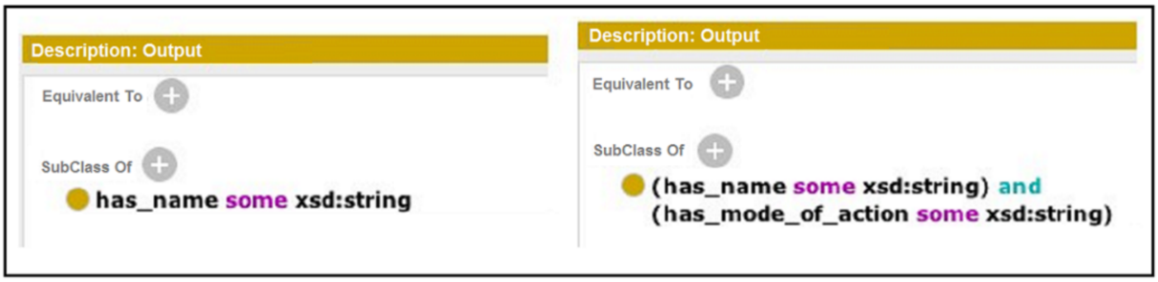
Domain Ontology

In the domain ontology, the data property *has_mode_of_action* is added. As a result, the service ontology becomes compatible and supports the extended query, as it only uses existing concepts and properties.

Positional-Slotted Object-Applicative Semantic Mapping Rules

The PSOA rules are also modified to populate the newly added data property *has_mode_of_action* as Textbox 3 shows. The lines 22’-24’ are added in contrast to the original version in Figure 5. Valet SADI can rewrite the modified input and output descriptions with the help of the mapping rules, and generate the SQL query and the complete program code automatically to fetch answers from the source data. The services are thus rebuilt and redeployed in the service registry so that the modified query returns the correct answers. Figure 10 shows the updated status of the services. The service getNameByInsecticideId is now shown in green because the service was rebuilt by Valet SADI and redeployed in the service registry. The exact time the service was rebuilt is tabulated in the *Time of rebuild* column.

Figure 8. Old (left) and new (right) output description of the service getNameByInsecticideId.



has_name some xsd:string

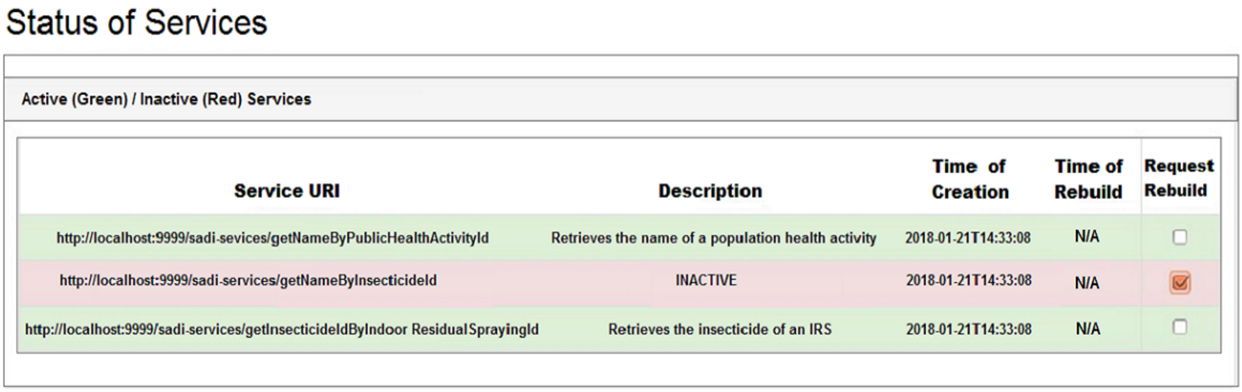
(has_name some xsd:string) and
(has_mode_of_action some xsd:string)

Table 1. Detection and identification of changes in the service ontology.

Timestamp	Description of change	Entity added	Entity deleted	Entity renamed	Affected service	Affected query
2018-01-21T14:33:08	An entity is added to the output definition	*has_mode_of_action*	N/A[a]	N/A	getNameByInsecticideId	Which indoor residual sprayings used permethrin as an insecticide?
2018-01-21T4:33:08	An entity is added to the output definition	xsd:string	N/A	N/A	getNameByInsecticideId	Which indoor residual sprayings used permethrin as an insecticide?

[a] N/A: not applicable.

Figure 9. Status of Semantic Automated Discovery and Integration (SADI) services. Active services are shown in green and inactive services in red.



Status of Services

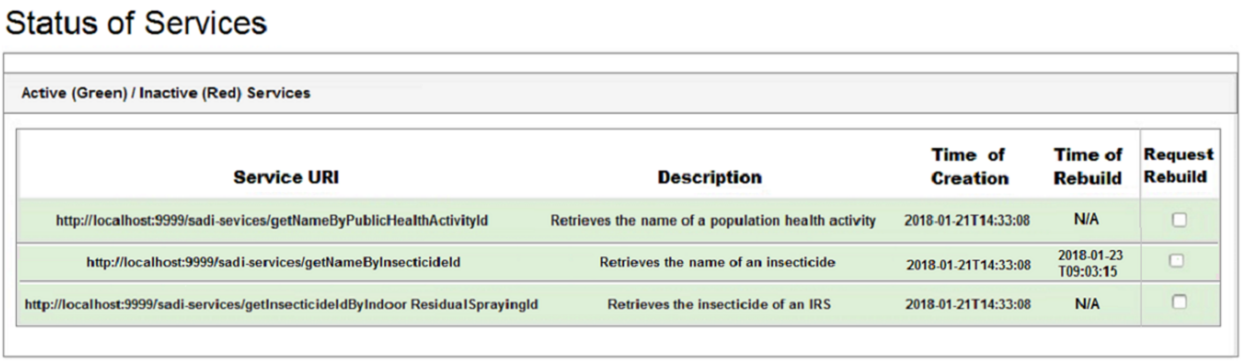
Active (Green) / Inactive (Red) Services

Service URI	Description	Time of Creation	Time of Rebuild	Request Rebuild
http://localhost:9999/sadi-sevices/getNameByPublicHealthActivityId	Retrieves the name of a population health activity	2018-01-21T14:33:08	N/A	☐
http://localhost:9999/sadi-services/getNameByInsecticideId	INACTIVE	2018-01-21T14:33:08	N/A	☑
http://localhost:9999/sadi-services/getInsecticideIdByIndoor ResidualSprayingId	Retrieves the insecticide of an IRS	2018-01-21T14:33:08	N/A	☐

Textbox 3. Fragment of the updated positional-slotted object-applicative file.

```
12' Group (
13'  Forall ?insecticideID (identityForInsecticideToinsecticideID(
14'                        identityForInsecticide(?insecticideID)) = ?insecticideID)
15'  Forall ?P (identityForInsecticide(identityForInsecticideToinsecticideID(?P)) = ?P)
16'  Forall ?id ?name ?mode.of.action (
17'   Insecticide(identityForInsecticide(?id)) :-
18'    db_insecticide(?id ?name ?mode.of.action))
19'  Forall ?id ?name ?mode.of.action (
20'   has_name(identityForInsecticide(?id) ?name) :-
21'    db_insecticide(?id ?name ?mode.of.action))
22'  Forall ?id ?name ?mode.of.action (
23'   has_mode_of_action(identityForInsecticide(?id) ?mode.of.action) :-
24'    db_insecticide(?id ?name ?mode.of.action))
25' )
```

Figure 10. Status of services after being rebuilt by Valet SADI.



Status of Services

Active (Green) / Inactive (Red) Services

Service URI	Description	Time of Creation	Time of Rebuild	Request Rebuild
http://localhost:9999/sadi-sevices/getNameByPublicHealthActivityId	Retrieves the name of a population health activity	2018-01-21T14:33:08	N/A	☐
http://localhost:9999/sadi-services/getNameByInsecticideId	Retrieves the name of an insecticide	2018-01-21T14:33:08	2018-01-23 T09:03:15	☐
http://localhost:9999/sadi-services/getInsecticideIdByIndoor ResidualSprayingId	Retrieves the insecticide of an IRS	2018-01-21T14:33:08	N/A	☐

Discussion

Surveillance remains a challenge for the malaria community, and many factors play a role in limiting access to relevant data resources for analysis and reuse [13]. We investigated the suitability of a solution to the challenges of system interoperability, distributed data access, semantic integration of data, and semantic support for query composition.

Semantics, Interoperability, and Evolution for Malaria Analytics

The SIEMA framework comprises several technologies and standards and is further customized to address the proposed targeted needs and interests of surveillance practitioners. The contribution takes three main directions. First, using SADI Web services allows for easy access to distributed data. This task is simplified further using Valet SADI, which enables a programmer to create services in an efficient and straightforward way. Second, due to the user interface features of the HYDRA query engine, SIEMA offers end users a more appealing way to build surveillance queries. HYDRA's ability to discover and call the services that are needed for a query permits the user to simply use the data as an abstract construct without having to look at its actual structure. Third, to make the system more robust and flexible, a dashboard has been introduced. The dashboard informs users when changes have occurred that render the services or queries inactive. This enables users to know which queries may no longer be reliable and to identify which parts of the service infrastructure must be rebuilt to restore it to its fully interoperable state. Deployed together, this combination of technologies offered by the SIEMA framework exhibits key functionalities that are of great value to the community.

Our initial studies in malaria surveillance [23,24] and other domains [30,48,49] showed us that this approach is a viable solution for enhancing interoperability. The successful extension of this methodology to the malaria use case appears promising based on the results from the initial implementation. The addition of a dashboard and Valet SADI extends the capabilities of the SADI framework to make it suitable for change detection, service restoration, and preservation of interoperability.

Evaluation

Whereas a systematic evaluation of the many components of this framework, individually and together, is beyond the scope of this initial study, we are aware that other malaria surveillance systems in sub–Saharan African countries have been reported and evaluated in part [12,50] according to attributes recommended by US Centers for Disease Control and Prevention [51]. Indeed, it would be valuable to make some direct system comparisons. Regrettably, few technical details of these systems have been reported, and their key features—such as their architecture, supported data types and data quality, use of data representation standards, capacity for change management, and degree of interoperability supported—have not been disclosed. Without such information, it is not possible to compare them with SIEMA. It is apparent, however, that the systems were designed for centralized data warehousing rather than seamless access to distributed data, and this is a significant distinction.

A brief assessment of SIEMA according to attributes recommended by Centers for Disease Control and Prevention [52], namely simplicity, data quality, flexibility, stability, and timeliness, can be made based on our initial experiments. To address the attribute of *simplicity*, the activity of creating surveillance queries can be assessed. Whereas writing syntactic queries (in SPARQL) for submission to a query client manually requires that a user have expertise with the query language, the HYDRA GUI can save users time and allow more complex surveillance queries to be composed by persons with less technical skill. Figure 5 shows a graph representation of a query built with the GUI, which is in turn translated to SPARQL automatically. To address the attribute of *data quality*, SIEMA's adoption of SADI ensures that the World Wide Web Consortium standard OWL and RDF[S] are used to describe and format every piece of data accessed through the framework. The attribute of *flexibility* is also met, since access to distributed data can be provisioned by using SADI services that are easily deployed in a registry, making them discoverable and readily usable for a variety of ad hoc queries. Likewise, the attributes of *stability* and *timeliness* of the system are met as the implemented dashboard tracks service uptime and reports failures immediately after detecting the faults. System rebuilds with Valet SADI ensure stability so that data can continue to be provided and the system stays operational if changes occur. In this way, interoperability is preserved and any service downtime is kept to a minimum.

Overall, we anticipate that the ongoing trials with the SIEMA framework will give the research and development team further insight into real-world requirements for interoperability and change management in malaria surveillance, leading to further improvements in adaptability and performance. Given the critical need for timely integration of distributed data from multiple heterogeneous sources in an efficient way, we hope to build cooperative partnerships between multiple disciplines, organizations, and sectors. In addition, insights gained from this research are likely transferable to a range of global surveillance projects.

Conclusion

We have demonstrated that authentic questions asked in malaria surveillance can be formalized as queries and mapped to a combination of Semantic Web services designed to deliver target data from distributed data sources. We have shown that using SIEMA and leveraging terminologies from community-developed ontologies offer flexibility both for integrating data and for easily composing queries. The developed infrastructure also offers a solution to the problem of change management, an important process for maintaining interoperability and integrity of an integrated surveillance system. Given that changes in the form of addition, renaming, and deletion of terminologies can frequently occur in the face of evolving system requirements, we introduced a change management dashboard. This makes it possible to identify important changes, report on the status of services as a consequence of changes, and offer users the option to rebuild inactive services. The dashboard and service reauthoring routines serve as an important vehicle to maintain system interoperability of mission-critical global surveillance programs. The infrastructure has been implemented and its relevance has been demonstrated with an authentic use case, with the goal of soliciting further requirements from the malaria analytics community. In future work, we will deploy SIEMA on live dynamic data sources.

Acknowledgments

This work was funded by the Bill and Melinda Gates Foundation (OPP ID # 1162018). A license for the use of HYDRA was provided by IPSNP Computing Inc.

References

1. World Health Organization. World malaria report 2017. Geneva, Switzerland: WHO; 2017. URL: http://apps.who.int/iris/bitstream/handle/10665/259492/9789241565523-eng.pdf [accessed 2018-06-05] [WebCite Cache ID 6zwui4w8o]
2. Wang H, Naghavi M, Allen C, Barber R, Bhutta Z, Carter A, GBD 2015 Mortality and Causes of Death Collaborators. . [doi: 10.1016/S0140-6736(16)31012-1] [Medline: 27733281]
3. World Health Organization. Regional Office for Africa. The health of the people: what works - the African Regional Health Report 2014. URL: http://apps.who.int/iris/bitstream/10665/137377/4/9789290232612.pdf [accessed 2018-02-16] [WebCite Cache ID 6xM7NkfUb]
4. Bridges DJ, Winters AM, Hamer DH. Malaria elimination: surveillance and response. Pathog Glob Health 2012 Aug;106(4):224-231 [FREE Full text] [doi: 10.1179/2047773212Y.0000000035] [Medline: 23265423]
5. Bennett A, Avanceña ALV, Wegbreit J, Cotter C, Roberts K, Gosling R. Engaging the private sector in malaria surveillance: a review of strategies and recommendations for elimination settings. Malar J 2017 Dec 14;16(1):252 [FREE Full text] [doi: 10.1186/s12936-017-1901-1] [Medline: 28615026]
6. Zheng J, Cade J, Brunk B, Roos D, Stoeckert C, Sullivan S, et al. Malaria study data integration and information retrieval based on OBO foundry ontologies. In: CEUR Work-shop Proceedings 1747. 2016 Presented at: International Conference on Biological Ontologies; Aug 1-4, 2016; Corvallis, Oregon, USA p. 1-4.
7. Mboera L, Rumisha S, Mlacha T, Mayala B, Bwana V, Shayo E. Malaria surveillance and use of evidence in planning and decision making in Kilosa district, Tanzania. Tanzan J Health Res 2017;19(3):1-10 [FREE Full text] [doi: 10.4314/thrb.v19i3.7]
8. Ohiri K, Ukoha NK, Nwangwu CW, Chima CC, Ogundeji YK, Rone A, et al. An assessment of data availability, quality, and use in malaria program decision making in Nigeria. Health Syst Reform 2016 Sep 23;2(4):319-330. [doi: 10.1080/23288604.2016.1234864]
9. Ohrt C, Roberts KW, Sturrock HJW, Wegbreit J, Lee BY, Gosling RD. Information systems to support surveillance for malaria elimination. Am J Trop Med Hyg 2015 Jul;93(1):145-152 [FREE Full text] [doi: 10.4269/ajtmh.14-0257] [Medline: 26013378]
10. Fu C, Lopes S, Mellor S, Aryal S, Sovannaroth S, Roca-Feltrer A. Experiences from developing and upgrading a web-based surveillance system for malaria elimination in Cambodia. JMIR Public Health Surveill 2017 Jun 14;3(2):e30 [FREE Full text] [doi: 10.2196/publichealth.6942] [Medline: 28615155]
11. Lu G, Liu Y, Beiersmann C, Feng Y, Cao J, Müller O. Challenges in and lessons learned during the implementation of the 1-3-7 malaria surveillance and response strategy in China: a qualitative study. Infect Dis Poverty 2016 Oct 05;5(1):94 [FREE Full text] [doi: 10.1186/s40249-016-0188-8] [Medline: 27716398]
12. Ibrahim B, Abubakar A, Bajoga U, Nguku P. Evaluation of the malaria surveillance system in Kaduna state, Nigeria 2016. Online J Public Health Inform 2017 May 02;9(1):e177. [doi: 10.5210/ojphi.v9i1.7775]
13. Briand D, Roux E, Desconnets JC, Gervet C, Barcellos C. From global action against malaria to local issues: state of the art and perspectives of web platforms dealing with malaria information. Malar J 2018 Mar 21;17(1):122 [FREE Full text] [doi: 10.1186/s12936-018-2270-0] [Medline: 29562918]
14. World Health Organization. Global technical strategy for malaria 2016-2030. Geneva, Switzerland: WHO; 2015. URL: http://apps.who.int/iris/bitstream/handle/10665/176712/9789241564991_eng.pdf [accessed 2018-06-04] [WebCite Cache ID 6zvoZNpiT]
15. The FAIR Data Principle. La Jolla, CA: FORCE11; 2017. URL: https://www.force11.org/group/fairgroup/fairprinciples [accessed 2018-02-19] [WebCite Cache ID 6xOqTBuRc]
16. Semantic Web. Cambridge, MA: W3C; 2013. URL: http://www.w3.org/2001/sw/ [accessed 2018-04-09] [WebCite Cache ID 6yYfCTxtC]
17. The Malaria Atlas Project. Oxford, UK: Malaria Atlas Project; 2018. URL: https://map.ox.ac.uk [accessed 2018-02-19] [WebCite Cache ID 6yZrkKnap]
18. Africa Development Indicators 2012/13. Washington, DC: World Bank; 2013. URL: https://openknowledge.worldbank.org/bitstream/handle/10986/13504/9780821396162.pdf [accessed 2018-06-04] [WebCite Cache ID 6zvpE1xH6]

19. Ontologies: vocabularies. Cambridge, MA: W3C; 2015. URL: https://www.w3.org/standards/semanticweb/ontology [accessed 2017-07-02] [WebCite Cache ID 6zvqCphpN]
20. Lozano-Fuentes S, Bandyopadhyay A, Cowell LG, Goldfain A, Eisen L. Ontology for vector surveillance and management. J Med Entomol 2013 Jan;50(1):1-14 [FREE Full text] [Medline: 23427646]
21. Topalis P, Mitraka E, Dritsou V, Dialynas E, Louis C. IDOMAL: the malaria ontology revisited. J Biomed Semantics 2013 Sep 13;4(1):16 [FREE Full text] [doi: 10.1186/2041-1480-4-16] [Medline: 24034841]
22. Dialynas E, Topalis P, Vontas J, Louis C. MIRO and IRbase: IT tools for the epidemiological monitoring of insecticide resistance in mosquito disease vectors. PLoS Negl Trop Dis 2009 Jun 23;3(6):e465 [FREE Full text] [doi: 10.1371/journal.pntd.0000465] [Medline: 19547750]
23. Brenas JH, Al-Manir MS, Baker CJO, Shaban-Nejad A. A malaria analytics framework to support evolution and interoperability of global health surveillance systems. IEEE Access 2017;5:21605-21619 [FREE Full text] [doi: 10.1109/ACCESS.2017.2761232]
24. Brenas JH, Al Manir MS, Zinszer K, Baker CJO, Shaban-Nejad A. Exploring semantic data federation to enable malaria surveillance queries. Stud Health Technol Inform 2018;247:6-10. [doi: 10.3233/978-1-61499-852-5-6] [Medline: 29677912]
25. Shaban-Nejad A, Haarslev V. Managing changes in distributed biomedical ontologies using hierarchical distributed graph transformation. Int J Data Min Bioinform 2015;11(1):53-83. [doi: 10.1504/IJDMB.2015.066334] [Medline: 26255376]
26. Wilkinson MD, Vandervalk B, McCarthy L. The semantic automated discovery and integration (SADI) web service design-pattern, API and reference implementation. J Biomed Semantics 2011 Oct 24;2(1):8 [FREE Full text] [doi: 10.1186/2041-1480-2-8] [Medline: 22024447]
27. Brickley D, Guha R. RDF schema 1.1. Cambridge, MA: W3C; 2014 Feb 25. URL: https://www.w3.org/TR/rdf-schema/ [accessed 2018-02-19] [WebCite Cache ID 6xMBgRnRL]
28. OWL Working Group. Web Ontology Language (OWL). Cambridge, MA: W3C; 2013 Dec 11. URL: https://www.w3.org/OWL/ [accessed 2018-02-19] [WebCite Cache ID 6xMBmllUX]
29. Vandervalk P, McCarthy E, Wilkinson M. SHARE: a Semantic Web query engine for bioinformatics. 2009 Presented at: The Semantic Web, Fourth Asian Conference, ASWC 2009; Dec 6-9, 2009; Shanghai, China p. 367-369. [doi: 10.1007/978-3-642-10871-6_27]
30. Riazanov A, Klein A, Shaban-Nejad A, Rose GW, Forster AJ, Buckeridge DL, et al. Semantic querying of relational data for clinical intelligence: a semantic web services-based approach. J Biomed Semantics 2013;4(1):9 [FREE Full text] [doi: 10.1186/2041-1480-4-9] [Medline: 23497556]
31. HYDRA. Saint John, NB: IPSNP Computing Inc URL: http://ipsnp.com/HYDRA/ [accessed 2018-04-10] [WebCite Cache ID 6yZvQ1UZD]
32. Zinszer K, Shaban-Nejad A, Menon S, Okhmatovskaia A, Carroll L, Painter I, et al. Integrated disease surveillance to reduce data fragmentation - an application to malaria control. Online J Public Health Inform 2015;7(1):e181. [doi: 10.5210/ojphi.v7i1.5849]
33. World Health Organization. Global Malaria Mapper. Geneva, Switzerland: WHO; 2018. URL: http://www.who.int/malaria/publications/world_malaria_report/global_malaria_mapper/en/ [accessed 2018-01-17] [WebCite Cache ID 6xMCOBAmb]
34. Okhmatovskaia A, Shaban-Nejad A, Lavigne M, Buckeridge D. Addressing the challenge of encoding causal epidemiological knowledge in formal ontologies: a practical perspective. Stud Health Technol Inform 2014;205:1125-1129. [Medline: 25160364]
35. CDISC. Malaria Therapeutic Area User Guide v1.0. Austin, TX: Clinical Data Interchange Standards Consortium; 2017 Jan 09. URL: https://www.cdisc.org/standards/therapeutic-areas/malaria [accessed 2018-04-12] [WebCite Cache ID 6yd6sJ5Iv]
36. W3C SPARQL Working Group. SPARQL 1.1 overview. Cambridge, MA: W3C; 2013 Mar 21. URL: https://www.w3.org/TR/sparql11-overview/ [accessed 2018-02-01] [WebCite Cache ID 6xMDu2OWw]
37. Huntley RP, Sawford T, Martin MJ, O'Donovan C. Understanding how and why the Gene Ontology and its annotations evolve: the GO within UniProt. Gigascience 2014 Mar 18;3(1):4 [FREE Full text] [doi: 10.1186/2047-217X-3-4] [Medline: 24641996]
38. Kondylakis H, Flouris G, Plexousakis D. Ontology and schema evolution in data integration: review and assessment. 2009 Presented at: On the Move to Meaningful Internet Systems: OTM 2009, Confederated International Conferences, CoopIS, DOA, IS,ODBASE 2009; Nov 1-6, 2009; Vilamoura, Portugal p. 932-947. [doi: 10.1007/978-3-642-05151-7_14]
39. Brenas J, Al Manir MS, Baker C, Shaban-Nejad A. A change management dashboard for the SIEMA global malaria surveillance infrastructure. 2017 Presented at: ISWC Posters & Demonstrations and Industry Tracks co-located with 16th International Semantic Web Conference (ISWC 2017); Oct 23-25, 2017; Vienna, Austria.
40. Lambrix P, Dragisic Z, Ivanova V, Anslow C. Visualization for ontology evolution. 2016 Presented at: Second International Workshop on Visualization and Interaction for Ontologies and Linked Data co-located with the 15th International Semantic Web Conference, VOILA@ISWC 2016; Oct 17, 2016; Kobe, Japan p. 54-67.
41. Flouris G, Manakanatas D, Kondylakis H, Plexousakis D, Antoniou G. Ontology change: classification and survey. Knowl Eng Rev 2008 May 23;23(02):117-152. [doi: 10.1017/S0269888908001367]

42. Meurice L, Cleve A. DAHLIA: a visual analyzer of database schema evolution. 2014 Presented at: 2014 Software Evolution Week - IEEE Conference on Software Maintenance, Reengineering, and Reverse Engineering, CSMR-WCRE; Feb 3-6, 2014; Antwerp, Belgium. [doi: 10.1109/CSMR-WCRE.2014.6747219]
43. Al Manir MS, Riazanov A, Boley H, Klein A, Baker C. Valet SADI: provisioning SADI web services for semantic querying of relational databases. 2016 Presented at: 20th International Database Engineering & Applications Symposium, IDEAS ’16; Jul 11-13, 2016; Montreal, QC, Canada p. 248-255. [doi: 10.1145/2938503.2938543]
44. Boley H. A RIF-Style Semantics for RuleML-Integrated Positional-Slotted, Object-Applicative Rules, Rule-Based Reasoning, Programming, and Applications. Berlin, Germany: Springer-Verlag; 2011:194-211.
45. Ministry of Health, Republic of Uganda. National Malaria Control Program. Kampala, Uganda: The Ministry; 2018. URL: http://health.go.ug/programs/national-malaria-control-program [accessed 2018-02-19] [WebCite Cache ID 6xMEpVOZg]
46. U.S. President's Malaria Initiative, Africa Indoor Residual Spraying Project. 2018 Jan 1. Rwanda expands IRS to refugee camp URL: http://www.africairs.net/2018/01/rwanda-expands-irs-to-refugee-camp/ [accessed 2018-02-21] [WebCite Cache ID 6xOrBpkET]
47. WHO Pesticide Evaluation Scheme. WHO recommended insecticides for indoor residual spraying against malaria vectors. Geneva, Switzerland: WHO; 2015 Mar 2. URL: http://www.who.int/whopes/Insecticides_IRS_2_March_2015.pdf?ua=1 [accessed 2018-02-19] [WebCite Cache ID 6xMEyzJKs]
48. Riazanov A, Laurila JB, Baker CJO. Deploying mutation impact text-mining software with the SADI Semantic Web Services framework. BMC Bioinformatics 2011;12 Suppl 4:S6 [FREE Full text] [doi: 10.1186/1471-2105-12-S4-S6] [Medline: 21992079]
49. Chepelev LL, Riazanov A, Kouznetsov A, Low HS, Dumontier M, Baker CJO. Prototype semantic infrastructure for automated small molecule classification and annotation in lipidomics. BMC Bioinformatics 2011 Jul 26;12:303 [FREE Full text] [doi: 10.1186/1471-2105-12-303] [Medline: 21791100]
50. Joseph A, Patrick N, Lawrence N, Lilian O, Olufemi A. Evaluation of malaria surveillance system in Ebonyi state, Nigeria, 2014. Ann Med Health Sci Res 2017;7:101-104 [FREE Full text]
51. German RR, Lee LM, Horan JM, Milstein RL, Pertowski CA, Waller MN, Guidelines Working Group Centers for Disease Control and Prevention (CDC). Updated guidelines for evaluating public health surveillance systems: recommendations from the Guidelines Working Group. MMWR Recomm Rep 2001 Jul 27;50(RR-13):1-35; quiz CE1. [Medline: 18634202]
52. Centers for Disease Control and Prevention. Overview of evaluating surveillance systems. Atlanta, GA: CDC; 2013. URL: https://www.cdc.gov/globalhealth/healthprotection/fetp/training_modules/22/overview-of-eval-survsys_field-g_final_09262013.pdf [accessed 2018-06-08] [WebCite Cache ID 701iXHU2G]

Abbreviations

GUI: graphical user interface
IDOMAL: Infectious Disease Ontology-Malaria
MIRO: Mosquito Insecticide Resistance Ontology
OWL: Web Ontology Language
PSOA: positional-slotted object-applicative
RDF[S]: resource description framework schema
SADI: Semantic Automated Discovery and Integration
SIEMA: Semantics, Interoperability, and Evolution for Malaria Analytics
SQL: Structured Query Language
VSMO: Vector Surveillance and Management Ontology

Edited by G Eysenbach; submitted 26.02.18; peer-reviewed by H Chen, I Peterson, JT Amenyo; comments to author 23.03.18; revised version received 16.04.18; accepted 08.05.18; published 15.06.18.

Please cite as:
Al Manir MS, Brenas JH, Baker CJO, Shaban-Nejad A
A Surveillance Infrastructure for Malaria Analytics: Provisioning Data Access and Preservation of Interoperability
JMIR Public Health Surveill 2018;4(2):e10218
URL: http://publichealth.jmir.org/2018/2/e10218/
doi:10.2196/10218
PMID:29907554

Original Paper

Improving Visualization of Female Breast Cancer Survival Estimates: Analysis Using Interactive Mapping Reports

Awatef Ahmed Ben Ramadan[1,2,3,4], MBBCh, MPH, PhD; Jeannette Jackson-Thompson[1,3,4], MSPH, PhD; Chester Lee Schmaltz[1,3], PhD

[1]Missouri Cancer Registry and Research Center, University of Missouri-Columbia, Columbia, MO, United States
[2]Department of Mathematics, Science and Informatics, Penfield College, Mercer University, Atlanta, GA, United States
[3]Department of Health Management and Informatics, School of Medicine, University of Missouri-Columbia, Columbia, MO, United States
[4]MU Informatics Institute, University of Missouri-Columbia, Columbia, MO, United States

Corresponding Author:
Awatef Ahmed Ben Ramadan, MBBCh, MPH, PhD
Department of Mathematics, Science and Informatics
Penfield College
Mercer University
AACC Building Suite 475 Office 479
3001 Mercer University Dr
Atlanta, GA, 30341
United States
Phone: 1 (678) 547 6240
Fax: 1 (678) 547 6076
Email: benramadan_aa@mercer.edu

Abstract

Background: The Missouri Cancer Registry collects population-based cancer incidence data on Missouri residents diagnosed with reportable malignant neoplasms. The Missouri Cancer Registry wanted to produce data that would be of interest to lawmakers as well as public health officials at the legislative district level on breast cancer, the most common non-skin cancer among females.

Objective: The aim was to measure and interactively visualize survival data of female breast cancer cases in the Missouri Cancer Registry.

Methods: Female breast cancer data were linked to Missouri death records and the Social Security Death Index. Unlinked female breast cancer cases were crossmatched to the National Death Index. Female breast cancer cases in subcounty senate districts were geocoded using TIGER/Line shapefiles to identify their district. A database was created and analyzed in SEER*Stat. Senatorial district maps were created using US Census Bureau's cartographic boundary files. The results were loaded with the cartographic data into InstantAtlas software to produce interactive mapping reports.

Results: Female breast cancer survival profiles of 5-year cause-specific survival percentages and 95% confidence intervals, displayed in tables and interactive maps, were created for all 34 senatorial districts. The maps visualized survival data by age, race, stage, and grade at diagnosis for the period from 2004 through 2010.

Conclusions: Linking cancer registry data to the National Death Index database improved accuracy of female breast cancer survival data in Missouri and this could positively impact cancer research and policy. The created survival mapping report could be very informative and usable by public health professionals, policy makers, at-risk women, and the public.

(JMIR Public Health Surveill 2018;4(2):e42) doi:10.2196/publichealth.8163

KEYWORDS

survival; female breast cancer; Missouri; cancer registry

Introduction

In the United States, it is estimated that 12% of women will be diagnosed with breast cancer at one stage of their lives [1]. Nationally, the estimated new cases of breast cancer were 14% of all new cancer cases and the estimated deaths from breast cancer were 7% of all cancer deaths in 2013 [2].

Traditionally, incidence and mortality rates have been presented in data tables, a format that is easily understood by epidemiologists and statisticians, but one that does not meet the needs of all potential users of the data. Data visualization is an alternative means of portraying the burden of breast cancer at various levels (eg, county, region, state).

There is a critical need to build accurate fact sheets in the form of interactive and dynamic map reports of the breast cancer burden at the substate level in Missouri. Several studies emphasize the efficiency and importance of matching National Death Index (NDI) data to cancer registry data to ensure high quality and accurate population-based cancer survival statistics [3-5]. We matched the registry breast cancer data to the Social Security Death Index (SSDI) and the NDI. This contribution will be significant because, with more complete data to analyze, we can accurately estimate survival data for the State of Missouri.

Numerous evidence-based studies have concluded that the use of geographic mapping software allows users to interact in a timely manner with the datasets and publish high-quality interactive reports [6-8]. The Web-based mapping systems' contribution is significant because these systems will enable users to visualize cancer data easily, and users can share this data with contributors in fields related to the visualized cancer. Distribution of geospatial health data could help public health leaders and decision makers in designing, developing, and adopting effective and efficient strategies and programs to improve public health outcomes targeting specific subpopulations within geographical areas [6-8].

A study by Koenig et al [9] recognized the impact of the interactive mapping visualization of health data on the public health field and health care-related laws and decisions. The study spotted the need for more interaction between mapmakers and the mapping reports' beneficiaries [9].

The Missouri General Assembly includes 34 senators, each representing one of Missouri's 34 districts. Every senate district included an annual average population of approximately 90,000 female residents (176,000 total residents) between 2004 and 2010 (study period). Most of the districts included whole counties. In high population density areas, including the Kansas City metropolitan area, Saint Louis metropolitan area, and the city of Springfield, district limits do not follow county boundaries [10,11].

We aim to measure the survival proportions of female breast cancer cases in the Missouri Cancer Registry database and to further analyze these survival data by stage and grade at diagnosis, by race, by age, and by senatorial district in Missouri for the period from 2004 through 2010. We also aim to visualize the survival data by Missouri state senatorial district by creating interactive mapping reports.

Methods

The study design was an observational longitudinal epidemiological study. The Missouri Cancer Registry and Research Center updated vital status of female breast cancer cases by linking with death records from the Missouri Department of Health and Senior Services and the SSDI [12]. We extracted female breast cancer cases (59,674 covering all years in the Missouri Cancer Registry database) without a known date and cause of death and submitted a formatted file containing required fields to the National Center for Health Statistics for NDI linkage [13]. The NDI staff returned the search results. We assessed the results to identify true matches. Partially matched records were reviewed manually using specific criteria (eg, possible typos, use of spouse's social security number, change of surname, use of compound names in a different order, use of nicknames). We then updated the database with the linkage results.

The female breast cancer cases in counties split by senate districts were loaded into Esri's ArcMap [14] with the Census Bureau's TIGER/Line Shapefiles [15] to determine their district based on their latitude and longitude. For this project, we used the State Senate districts that were defined by the redistricting following Census 2010 [16].

A database was created in SEER*Stat, a statistical software package for analyzing cancer data [17]; this database included cases diagnosed from 2004 through 2010 in which the tumor was the first reportable in situ or malignant tumor diagnosed in the woman's lifetime. This resulted in a total of 24,908 malignant cases for most of the survival calculations and an additional 5130 in situ cases included only in stage-specific survival calculations. The 5-year cause-specific survival proportions and their 95% confidence intervals were calculated for female breast cancer cases diagnosed from 2004 through 2010. Survival was measured in terms of cause-specific survival using the Surveillance, Epidemiology, and End Results (SEER) program's cause-specific death classification recode as the endpoint [18]. The 5-year female breast cancer survival was calculated by age, race, stage, and grade for each senate district. To protect patient confidentiality, we suppressed cells with small numbers, employing a commonly used threshold of five or fewer cases [19].

The US Census Bureau's cartographic boundary files were used to create maps showing 115 Missouri counties (including the City of St Louis—a county-equivalent entity) and 34 state senatorial districts [20]. Five-year survival statistics were loaded, along with cancer incidence and mortality data and the cartographic boundary files, into InstantAtlas software to produce interactive mapping reports that display our study's results [21]. The interactive reports included maps, graphs, and tables for each county and Missouri senatorial district as well as for 20 regions formed by aggregating senate districts by county boundaries. The senate district grouped to county boundaries were created because mortality data was not available at the subcounty level.

The years of female breast cancer diagnoses we chose for this study were from January 1, 2004 through December 31, 2010, with survival calculated by including follow-up through December 31, 2011. When this project was started, 2011 was the most recent year with complete survival follow-up for female breast cancer cases. The case selection criteria we used for survival excluded cases diagnosed in 2011 because a relatively large number of cases diagnosed in that year may have been reported too late to be included in the death linkages or even too late to be included in the Missouri Cancer Registry database. The beginning year of the case selection criteria—2004—was chosen such that relatively stable estimates could be obtained for a wide variety of demographic groups of interest while still covering a relatively recent set of years (7 years total).

We classified female breast cancer cases as “early stage” if the stage at diagnosis was in situ or localized according to the Derived SEER Summary Stage 2000 field [22]; “late-stage” female breast cancer cases included regional and distant cases. Low-grade female breast cancer cases involved grades I and II; high-grade female breast cancer cases included grades III and IV.

Results

The senatorial districts’ 5-year cause-specific survival proportions of female breast cancer were categorized, as shown in Tables 1-4, according to the following groupings: all malignant cases, cases younger than 50 years, cases 50 to 64 years, cases 65 years or older, white cases, African-American cases, early-stage (in situ and local) cases, late-stage (regional and distant) cases, low-grade cases, and high-grade cases. These tables include female breast cancer case counts and survival data for all 34 senatorial districts and Missouri and the 95% confidence intervals of the measured survival data for all the previously mentioned categories. Using these tables, the reader can compare every district to one another, as well as to the state’s survival proportion.

The reports we created displayed survival data results in two layouts: an “area profile” focused on displaying many indicators for one or a small number of selected districts along with results from statistical hypothesis testing (Figure 1) and a “double map” that displays two indicators simultaneously along with a district-level scatterplot (Figure 2). These reports include combined maps and statistical data. The area profile map displays a single map and presents many indicators for each senatorial district and compares each district’s results to the State of Missouri. The double map centers on assessing the statistical associations (correlation coefficient, R^2, and the simple linear regression equation) among the chosen survival indicators. The screenshots displayed in Figures 1 and 2 show the final formats of the interactive mapping reports we built at the Missouri Cancer Registry and Research Center to display Missouri female breast cancer survival data along with other incidence and mortality data [23,24].

Table 1. Five-year cause-specific female breast cancer survival across different age groups by state senatorial district, Missouri, 2004-2010.

Senatorial district	<50 years		50-64 years		≥65 years	
	Cases, n	Survival %[a] (95% CI)	Cases, n	Survival %[a] (95% CI)	Cases, n	Survival %[a] (95% CI)
1	181	91.6 (85.5-95.3)	302	90.7 (85.6-94.0)	428	85.8 (81.3-89.3)
2	192	92.9 (87.2-96.1)	224	89.9 (84.2-93.6)	198	79.2 (71.7-84.9)
3	128	77.2 (67.6-84.3)	251	88.8 (83.7-92.4)	272	84.0 (77.4-88.8)
4	163	89.5 (82.2-93.9)	300	86.9 (81.7-90.7)	435	76.3 (70.9-80.8)
5	167	81.8 (74.2-87.4)	230	86.3 (79.6-91.0)	231	76.0 (68.6-81.8)
6	160	85.6 (77.9-90.7)	234	86.6 (80.7-90.8)	314	81.0 (75.1-85.7)
7	147	86.0 (77.6-91.4)	269	86.9 (81.0-91.1)	264	83.8 (78.0-88.1)
8	181	87.2 (80.4-91.7)	261	86.5 (80.3-90.8)	208	86.7 (80.2-91.2)
9	158	69.0 (59.6-76.7)	273	81.8 (75.4-86.7)	288	72.0 (65.2-77.7)
10	147	90.5 (82.8-94.8)	261	86.2 (79.9-90.6)	267	83.8 (77.9-88.2)
11	130	84.7 (75.6-90.6)	240	84.4 (78.2-88.9)	326	82.6 (77.0-87.0)
12	152	91.1 (84.1-95.1)	254	84.7 (78.6-89.2)	300	80.8 (74.6-85.6)
13	207	85.3 (78.3-90.2)	290	84.0 (78.3-88.2)	356	79.6 (73.9-84.2)
14	212	80.6 (73.4-86.0)	319	87.0 (81.8-90.8)	283	83.1 (76.9-87.7)
15	240	92.5 (87.5-95.5)	368	90.6 (86.4-93.6)	403	85.1 (80.2-88.9)
16	139	87.5 (80.0-92.3)	237	84.3 (78.1-88.9)	306	82.6 (76.9-87.1)
17	173	85.6 (77.5-91.0)	258	92.3 (87.9-95.2)	268	83.6 (77.0-88.4)
18	155	85.4 (77.1-90.8)	259	83.4 (77.1-88.1)	388	78.1 (72.7-82.6)
19	167	87.1 (79.9-91.8)	238	87.2 (81.0-91.5)	206	83.4 (75.5-89.0)
20	155	90.6 (83.8-94.6)	251	86.4 (80.0-90.8)	260	82.4 (75.7-87.4)
21	137	82.7 (73.1-89.1)	250	84.9 (78.9-89.3)	296	80.0 (73.6-85.1)
22	138	87.0 (78.3-92.4)	215	90.7 (84.8-94.5)	184	80.9 (73.0-86.7)
23	181	90.2 (83.5-94.3)	304	91.2 (86.9-94.2)	267	88.7 (83.3-92.4)
24	188	91.0 (84.8-94.8)	371	91.0 (86.9-93.9)	448	85.2 (80.6-88.8)
25	142	82.9 (74.3-88.9)	274	81.7 (75.7-86.4)	327	79.1 (73.2-83.8)
26	204	88.4 (81.9-92.6)	338	89.5 (84.1-93.2)	327	79.4 (73.8-83.9)
27	151	86.6 (77.9-92.1)	257	84.8 (78.9-89.2)	318	79.7 (73.5-84.5)
28	144	84.5 (76.2-90.1)	294	85.5 (79.5-89.8)	352	82.4 (77.2-86.6)
29	117	85.0 (75.2-91.2)	250	90.7 (85.6-94.0)	349	81.1 (75.3-85.6)
30	145	82.0 (73.2-88.1)	232	85.8 (78.8-90.6)	341	81.6 (76.0-86.0)
31	146	87.3 (78.6-92.6)	253	87.6 (81.5-91.8)	293	78.6 (72.0-83.8)
32	155	77.2 (67.8-84.1)	245	86.7 (80.7-91.0)	287	81.7 (75.8-86.4)
33	137	88.2 (80.2-93.1)	231	83.8 (77.4-88.6)	273	78.6 (71.9-83.8)
34	157	87.5 (80.1-92.3)	252	85.3 (78.7-90.0)	264	80.7 (74.4-85.6)
Missouri	5496	86.1 (85.0-87.2)	9085	87.0 (86.1-87.8)	10,327	81.4 (80.4-82.3)

[a] Five-year cause-specific survival.

Table 2. Five-year cause-specific female breast cancer survival data among whites and African Americans by state senatorial district, Missouri, 2004-2010.

Senatorial District	White		African American	
	Cases, n	Survival %[a] (95% CI)	Cases, n	Survival %[a] (95% CI)
1	871	88.8 (86.0-91.1)	27	88.4 (68.2-96.1)
2	590	87.1 (83.6-89.9)	12	88.9 (43.3-98.4)
3	643	84.7 (81.1-87.8)	—[b]	—
4	578	85.4 (81.5-88.5)	302	76.4 (70.2-81.5)
5	248	85.8 (80.2-90.0)	370	78.6 (73.0-83.1)
6	691	84.6 (81.1-87.4)	11	58.3 (23.0-82.1)
7	536	87.6 (84.0-90.5)	133	77.7 (67.1-85.3)
8	618	86.9 (83.3-89.8)	25	77.1 (53.2-89.8)
9	292	81.6 (75.7-86.2)	418	70.4 (64.6-75.5)
10	645	86.1 (82.5-89.0)	26	91.8 (71.1-97.9)
11	654	83.5 (79.9-86.6)	33	84.7 (58.4-95.0)
12	698	84.5 (80.9-87.4)	—	—
13	489	85.4 (81.3-88.7)	352	77.6 (71.7-82.5)
14	368	89.1 (84.6-92.4)	435	79.6 (74.6-83.7)
15	966	89.3 (86.7-91.4)	14	68.4 (35.9-86.8)
16	661	84.7 (81.2-87.6)	12	67.3 (27.7-88.5)
17	673	87.8 (84.5-90.5)	16	75.7 (41.6-91.6)
18	769	81.8 (78.3-84.8)	28	67.5 (41.8-83.8)
19	549	85.1 (81.1-88.4)	42	92.1 (77.5-97.4)
20	662	85.8 (82.2-88.8)	—	—
21	661	82.7 (78.8-85.9)	18	65.6 (34.3-84.7)
22	525	86.0 (82.0-89.2)	—	—
23	722	90.0 (87.1-92.3)	24	85.2 (60.6-95.0)
24	897	89.0 (86.2-91.2)	72	80.8 (67.4-89.2)
25	688	81.6 (77.9-84.7)	49	67.4 (48.1-80.8)
26	840	85.5 (82.3-88.1)	7	100.0 (—)
27	704	83.7 (80.1-86.7)	20	49.6 (14.6-77.4)
28	778	83.9 (80.5-86.7)	—	—
29	706	85.3 (81.8-88.1)	—	—
30	698	83.3 (79.6-86.3)	10	56.4 (7.5-88.1)
31	671	84.1 (80.3-87.2)	12	69.1 (29.4-89.4)
32	675	82.2 (78.4-85.4)	—	—
33	636	82.4 (78.6-85.6)	—	—
34	637	84.5 (80.8-87.5)	19	72.8 (41.2-89.2)
Missouri	22,039	85.4 (84.8-85.9)	1607	76.8 (74.7-78.7)

[a] Five-year cause-specific survival.

[b] "—" indicates Survival statistics suppressed due to five or fewer cases.

Table 3. Five-year cause-specific female breast cancer survival data by stage at diagnosis and state senatorial district, Missouri, 2004-2010.

Senatorial district	All malignant cases (excludes in situ but includes unstaged cases)		Early stage (in situ & local)		Late stage (regional & distant)	
	Cases, n	Survival %[a] (95% CI)	Cases, n	Survival %[a] (95% CI)	Cases, n	Survival %[a] (95% CI)
1	911	88.6 (85.9-90.9)	763	96.8 (94.7-98.1)	328	77.5 (71.3-82.6)
2	614	87.3 (83.9-90.1)	497	97.5 (95.4-98.6)	244	75.5 (68.5-81.2)
3	651	84.4 (80.7-87.4)	467	96.1 (93.2-97.8)	267	72.3 (65.3-78.2)
4	898	82.5 (79.2-85.2)	703	95.4 (93.0-97.0)	365	68.6 (62.5-73.9)
5	628	81.5 (77.6-84.8)	459	95.1 (92.0-97.0)	274	66.3 (58.9-72.6)
6	708	84.0 (80.5-86.9)	637	94.7 (92.2-96.4)	232	70.7 (63.5-76.7)
7	680	85.4 (81.9-88.3)	562	95.9 (93.3-97.5)	283	74.4 (67.8-79.8)
8	650	86.7 (83.2-89.5)	555	95.5 (92.8-97.2)	254	78.0 (71.0-83.5)
9	719	75.0 (70.9-78.6)	550	92.1 (88.7-94.5)	323	60.4 (53.5-66.6)
10	675	86.2 (82.7-89.0)	536	95.7 (93.1-97.3)	251	76.4 (68.9-82.2)
11	696	83.6 (80.1-86.6)	547	94.6 (91.8-96.4)	272	74.1 (67.5-79.6)
12	706	84.5 (80.9-87.4)	552	95.3 (92.6-97.0)	269	71.7 (64.5-77.8)
13	853	82.6 (79.3-85.4)	662	94.6 (92.1-96.3)	355	69.8 (63.7-75.1)
14	814	83.9 (80.6-86.7)	627	96.2 (93.7-97.7)	362	71.7 (65.8-76.8)
15	1011	89.0 (86.4-91.1)	889	96.3 (94.3-97.6)	344	78.3 (72.6-82.9)
16	682	84.3 (80.8-87.2)	529	95.2 (92.5-96.9)	256	74.5 (67.7-80.2)
17	699	87.4 (84.1-90.1)	549	97.7 (95.5-98.9)	261	73.4 (66.1-79.3)
18	802	81.3 (77.8-84.3)	608	94.9 (92.2-96.7)	309	67.7 (61.1-73.4)
19	611	86.0 (82.3-89.0)	497	95.7 (92.6-97.5)	230	73.0 (65.5-79.0)
20	666	85.9 (82.3-88.8)	576	95.6 (93.0-97.3)	240	77.1 (70.0-82.7)
21	683	82.3 (78.5-85.5)	486	95.8 (92.8-97.6)	296	68.7 (61.7-74.7)
22	537	86.4 (82.5-89.5)	441	95.4 (92.4-97.2)	191	76.2 (68.1-82.6)
23	752	90.0 (87.2-92.2)	616	97.8 (95.6-98.9)	293	80.4 (74.5-85.1)
24	1007	88.5 (85.9-90.7)	854	96.5 (94.5-97.8)	343	77.2 (71.3-81.9)
25	743	80.8 (77.2-83.9)	516	94.7 (92.0-96.6)	291	66.4 (59.4-72.5)
26	869	85.4 (82.3-88.1)	733	94.9 (92.4-96.6)	302	74.1 (67.6-79.5)
27	726	83.0 (79.4-86.0)	549	94.6 (91.7-96.5)	272	69.6 (62.3-75.7)
28	790	84.0 (80.6-86.8)	603	95.7 (93.1-97.4)	301	73.3 (67.0-78.6)
29	716	85.2 (81.8-88.0)	571	96.7 (94.4-98.1)	259	71.5 (63.8-77.9)
30	718	83.0 (79.4-86.1)	611	94.2 (91.2-96.2)	242	69.5 (62.1-75.7)
31	692	83.8 (80.1-86.9)	555	94.4 (91.4-96.4)	256	75.0 (67.6-80.9)
32	687	82.4 (78.7-85.6)	550	94.7 (91.8-96.5)	250	67.2 (59.5-73.8)
33	641	82.6 (78.7-85.7)	542	93.0 (89.6-95.3)	205	69.7 (62.0-76.2)
34	673	84.1 (80.5-87.1)	543	96.0 (93.4-97.6)	251	70.8 (63.3-76.9)
Missouri	24,908	84.5 (84.0-85.0)	19,935	95.5 (95.1-95.8)	9471	72.3 (71.2-73.4)

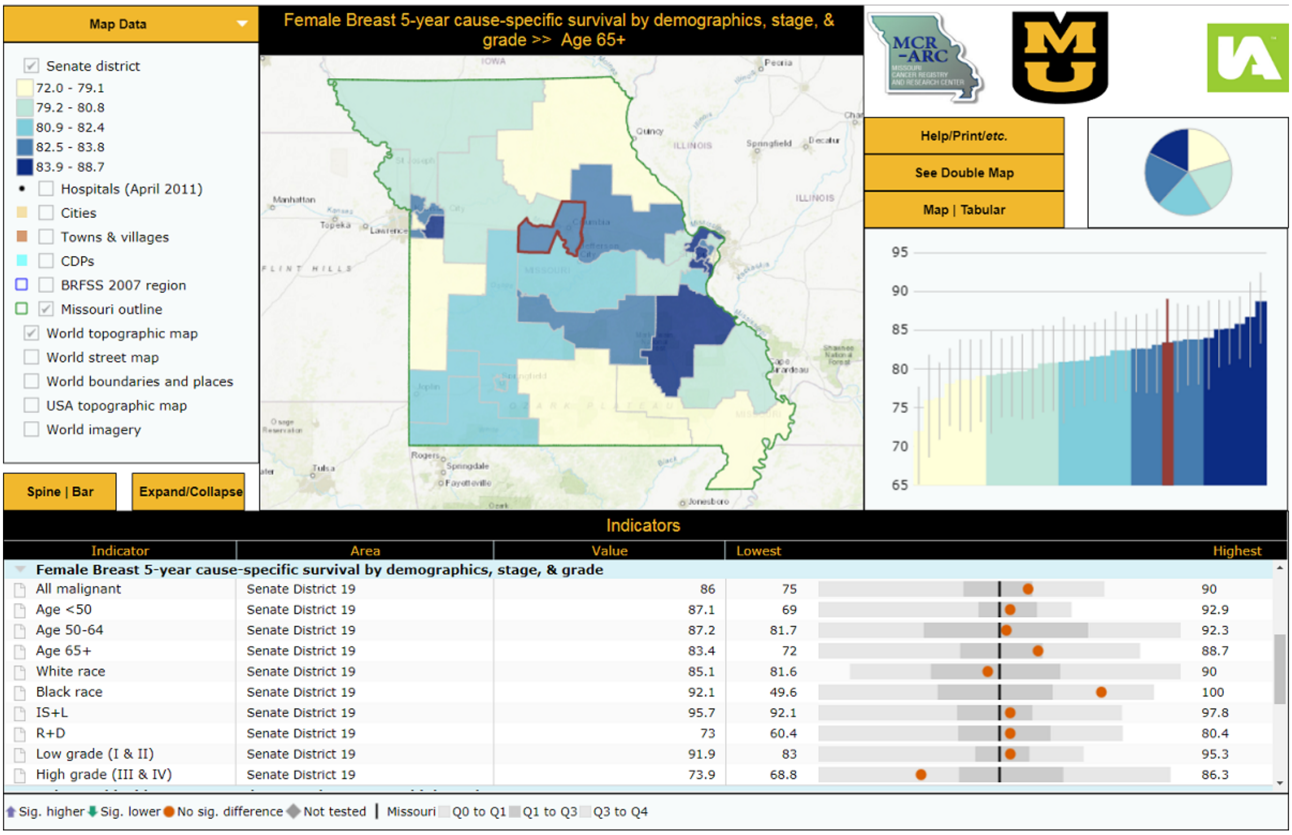
[a] Five-year cause-specific survival.

Table 4. Five-year cause-specific female breast cancer survival data for low- and high-grade cases, Missouri, 2004-2010.

Senatorial district	Low-grade (I & II)		High-grade (III & IV)	
	Cases, n	Survival %[a] (95% CI)	Cases, n	Survival %[a] (95% CI)
1	576	93.8 (90.6-95.9)	232	84.0 (77.9-88.6)
2	395	94.2 (90.8-96.4)	184	81.5 (74.0-87.0)
3	334	93.2 (89.1-95.8)	239	75.9 (68.5-81.8)
4	510	90.5 (86.8-93.2)	321	75.8 (69.5-80.9)
5	309	87.6 (82.3-91.3)	273	76.8 (70.3-82.1)
6	431	90.8 (86.9-93.6)	205	77.3 (70.0-83.1)
7	420	93.1 (89.4-95.5)	208	76.9 (69.3-82.8)
8	405	91.9 (87.6-94.8)	200	77.6 (70.2-83.5)
9	416	83.0 (77.8-87.1)	247	68.8 (61.5-75.0)
10	444	91.1 (86.8-94.0)	172	79.7 (71.7-85.7)
11	431	90.3 (86.6-93.0)	202	75.7 (68.0-81.8)
12	443	91.2 (87.5-93.9)	203	76.0 (67.5-82.6)
13	494	89.9 (86.1-92.7)	296	71.3 (64.8-76.8)
14	433	91.9 (88.1-94.5)	325	77.7 (71.7-82.5)
15	605	95.2 (92.5-96.9)	308	81.7 (75.7-86.3)
16	383	90.0 (85.8-93.0)	240	80.5 (73.9-85.5)
17	454	91.5 (87.9-94.1)	202	82.4 (74.4-88.2)
18	474	90.2 (86.3-93.0)	233	75.7 (68.5-81.5)
19	378	91.9 (87.6-94.8)	197	73.9 (65.8-80.3)
20	377	95.3 (92.0-97.3)	257	77.7 (70.9-83.1)
21	427	89.6 (85.4-92.7)	199	76.8 (69.1-82.8)
22	290	94.4 (89.8-96.9)	200	77.6 (69.7-83.7)
23	501	92.3 (88.9-94.8)	220	86.3 (80.2-90.7)
24	634	92.8 (89.8-95.0)	292	85.3 (79.8-89.4)
25	386	89.1 (84.4-92.4)	253	73.7 (66.9-79.4)
26	539	91.9 (88.4-94.3)	271	77.8 (71.0-83.2)
27	402	90.8 (86.7-93.7)	250	76.8 (69.7-82.4)
28	442	91.8 (88.0-94.4)	289	81.9 (76.0-86.5)
29	428	91.5 (87.4-94.3)	243	81.2 (74.7-86.3)
30	402	91.4 (87.4-94.1)	281	77.0 (70.1-82.6)
31	393	90.9 (86.6-93.9)	235	80.9 (73.6-86.3)
32	368	92.8 (88.6-95.5)	277	72.1 (65.1-78.0)
33	358	92.0 (87.8-94.8)	240	75.4 (68.4-81.1)
34	463	89.4 (85.4-92.3)	159	78.7 (69.2-85.5)
Missouri	14,745	91.4 (90.8-92.0)	8153	77.8 (76.7-78.8)

[a] Five-year cause-specific survival.

Figure 1. Area Profile Interactive Report Displaying FBC 5-Year Cause-specific Survival Data by Senatorial District [23].

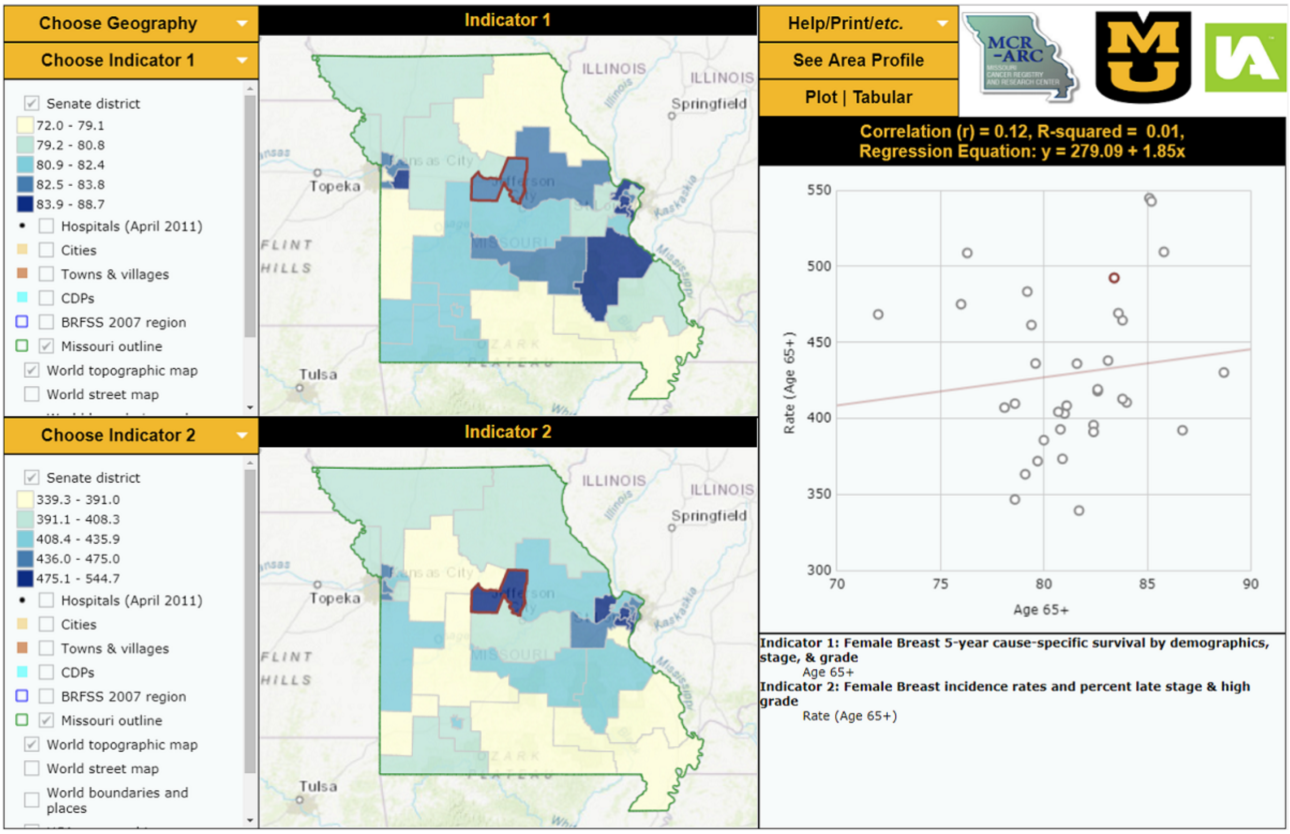


Female Breast 5-year cause-specific survival by demographics, stage, & grade >> Age 65+

Indicators

Indicator	Area	Value	Lowest	Highest
Female Breast 5-year cause-specific survival by demographics, stage, & grade				
All malignant	Senate District 19	86	75	90
Age <50	Senate District 19	87.1	69	92.9
Age 50-64	Senate District 19	87.2	81.7	92.3
Age 65+	Senate District 19	83.4	72	88.7
White race	Senate District 19	85.1	81.6	90
Black race	Senate District 19	92.1	49.6	100
IS+L	Senate District 19	95.7	92.1	97.8
R+D	Senate District 19	73	60.4	80.4
Low grade (I & II)	Senate District 19	91.9	83	95.3
High grade (III & IV)	Senate District 19	73.9	68.8	86.3

Figure 2. Double Map Interactive Report Displaying FBC 5-Year Cause-specific Survival Data by Senatorial District [24].



Correlation (r) = 0.12, R-squared = 0.01, Regression Equation: y = 279.09 + 1.85x

Indicator 1: Female Breast 5-year cause-specific survival by demographics, stage, & grade
Age 65+
Indicator 2: Female Breast incidence rates and percent late stage & high grade
Rate (Age 65+)

Discussion

Principal Findings

The Missouri Cancer Registry needs to measure female breast cancer survival proportions to be able to evaluate the impact of Missouri's breast cancer control program and the burden of female breast cancer in Missouri. The measured and visualized survival data will transform our registry from being an incidence registry to becoming a survival registry for breast cancer.

Survival data mirrors female breast cancer prediction in a specific period [25]. We used the Missouri Cancer Registry records because Missouri Cancer Registry is a nationally recognized, population-based registry with data that originates from diverse sources including hospitals, ambulatory surgical centers, freestanding cancer treatment centers, pathology laboratories, long-term care facilities, and physician offices. It also contains cases obtained through case-sharing agreements with 19 states. The Missouri Cancer Registry data undergo a strict quality control process and the data are evaluated following specific national measures [26]. Several studies have revealed the significance of linking NDI data to a central cancer registry's data to obtain more accurate population-based cancer survival data [3-5].

From this study's results, as shown in Tables 1-4, we can create female breast cancer survival profiles for the 34 Missouri senate districts. By creating these profiles, we can compare each district's results to the state and to other districts' results and give more detailed information to public health practitioners and decision makers about female breast cancer in their district.

Mapping Reports

Cancer incidence and mortality data have traditionally been presented in tabular and descriptive statistics formats; these are easily understood by health professionals with specific knowledge and experience in statistics and epidemiology. At the Missouri Cancer Registry and Research Center, we strive to present our data in formats that meet the needs of a wide range of potential data users. That is why we chose to combine our survival data with geographical data to produce interactive mapping reports at the Missouri senate district level. InstantAtlas is an interactive, internet-based mapping tool licensed to the Missouri Cancer Registry and Research Center that allows users to visually display data gathered from the registry and other databases. Use of interactive data visualization and mapping software allows users to interact with the datasets. We built two interactive mapping reports that include our senate district-level female breast cancer survival data [23,24]. The two maps, the area profile map and the double map, have not yet been published on the Missouri Cancer Registry and Research Center's website. The area profile report shows a single map and focuses on displaying many indicators for a selected state senate district and compares the district's findings to other districts and to Missouri. The double map focuses on exploring the relationships between selected indicators; it displays two indicators simultaneously along with a scatterplot or a table.

The InstantAtlas reports can facilitate communication between collaborators from different fields related to breast cancer, enhance female breast cancer research and policy, and inform public health professionals and policy makers. These maps can be used as educational tools at the community level for women at risk and the public about the distribution of female breast cancer in Missouri by age, race, stage and grade at diagnosis, and by senatorial district. These data could be used as a knowledge base at Missouri oncology facilities to assess management plan decisions taken by providers and by female breast cancer cases.

Study Challenges and Limitations

During the matching processes, some cases did not have a social security number, which is the best available unique identifier. Also, some identifiers, such as date of birth and last and/or first name, showed differences when the NDI database and the registry database were compared, possibly due to data entry errors or changed last name. Such cases were manually reviewed. Manual review of all partial matches was done by more than one Missouri Cancer Registry and Research Center staff member, including at least one certified tumor registrar, to reduce possible mistakes.

Survival was measured using cause-specific survival rather than relative survival (another common net measure of survival) to avoid the need of having detailed population lifetables by senatorial district. Potential disadvantages of using cause-specific survival is that, unlike relative survival, it relies on additionally having the cause of death rather than just the fact and date of death and on accurate coding of the cause of death [18]. To decrease the number of known decedents with unknown cause of death in the Missouri Cancer Registry database, these cases were included in the NDI linkage to try to obtain their cause of death. To lessen the impact of miscoded cause of death (eg, a breast cancer death being misattributed to the location of a metastatic site), the data used here was defined "breast cancer death" according to the SEER cause-specific death classification recode variable [18]. It should be noted that this will miss indirect deaths originating from a diagnosis of breast cancer, such as toxic effects of chemotherapy. Moreover, the use of this death classification variable is limited to first primary tumors only as used in these analyses and cannot be used to analyze second and subsequent tumors.

Due to aggregating the cases to areal units, this study is subject to the modifiable areal unit problem [27]. State senate areal units were selected for this project because they would be relevant to policy makers making decisions at the senate district level and to constituents within those districts. It should be noted that the modifiable areal unit problem implies that differing conclusions can potentially be drawn from the same data had different areal units been used.

The survival rates presented in these mapping reports are the observed percentages rather than rates that have been spatially smoothed. Observed percentages may be more directly interpretable and relevant to the residents of each of the individual senate districts; however, observed percentages have the disadvantage of being less stable and more prone to spurious high and low values than spatially smoothed survival rates. This instability is mitigated somewhat by the fact that, with the notable exception of African Americans in many districts, survival was calculated with fairly large sample sizes: always more than 100 and generally at least 200 or more.

For the selected cases, only approximately 1% had the district imputed. Due to the relatively small number of cases, a sensitivity analysis was not performed.

Future Directions

In the future, by combining mortality and incidence data in the survival profiles, we will be able to inform every district's decision makers about the full picture of female breast cancer burden by district and we could help them assess female breast cancer interventions and policies on geographical bases. Due to small sample sizes, we do not have county-level results from the Behavioral Risk Factor Surveillance System, a state-based health survey that annually gathers data on health events, behaviors, preventive practices, and access to health care. A similar Behavioral Risk Factor Surveillance System-based survey known as the "County-Level Study" has been conducted at the county level in Missouri [28]. In the future, we hope to combine these results with female breast cancer survival data and create InstantAtlas mapping reports at the senatorial district level that include survival and other measured contextual indicators (eg , demographic, environment, and socioeconomic), similar to the currently published county-level maps. This kind of mapping report could be used to explore the relationship between female breast cancer and other measured contextual indicators all over Missouri.

In this paper, we measured 5-year cause-specific survival proportions of female breast cancer for the 34 senate districts in Missouri. In the future, we will consider the feasibility of measuring the same data for all 163 Missouri legislative districts [29,30]. We will also consider measuring 5-year cause-specific survival for other screening-amenable cancers (eg, colorectal cancers) and for cancers that impact many residents (eg, lung cancer).

Before we publish senate district maps on our website, we aim to test the usability of the survival maps using a pilot sample of actual users, similar to one we conducted with our previously published maps [31], in order to make them more user friendly.

Conclusions

Net measures of survival factor out other causes of death and are useful from a policy-based perspective. These measures enable comparisons of cancer survival across geographical regions and between groups of patients without differences in background mortality rates of other causes impacting the results.

Cancer registry data are very rich and can be used in the exploration of many scientific theories and models. Registry data are a valuable source for survival data on breast cancer by race, age, and stage at diagnosis. Using cancer registry data supplemented by SSDI and NDI information will be beneficial and can improve accuracy of breast cancer survival data by age, stage, or race, as well as by geographic area (counties and senatorial districts).

Acknowledgments

The authors would like to thank the anonymous reviewers of the first version of this paper for their comprehensive and valuable feedback resulting in an extensive refocusing and reorganization of this paper. We would also like to thank the reporting facility and central registry staff for their efforts.

Core activities of the Missouri Cancer Registry are supported in part by a cooperative agreement between the Centers for Disease Control and Prevention and the Missouri Department of Health and Senior Services (DHSS) (5NU58DP003924-04/05) and a Surveillance Contract between DHSS and the University of Missouri. The publication's content is exclusively the responsibility of the authors and does not necessarily reflect the views of the funding agencies.

Conflicts of Interest

None declared.

References

1. DeSantis C, Ma J, Bryan L, Jemal A. Breast cancer statistics, 2013. CA Cancer J Clin 2014;64(1):52-62 [FREE Full text] [doi: 10.3322/caac.21203] [Medline: 24114568]
2. National Cancer Institute. 2014. Cancer Stat Facts: Female Breast Cancer URL: http://seer.cancer.gov/statfacts/html/breast.html [WebCite Cache ID 6nGmpKo1l]
3. Johnson CJ, Weir HK, Fink AK, German RR, Finch JL, Rycroft RK, Accuracy of Cancer Mortality Study Group. The impact of National Death Index linkages on population-based cancer survival rates in the United States. Cancer Epidemiol 2013 Feb;37(1):20-28 [FREE Full text] [doi: 10.1016/j.canep.2012.08.007] [Medline: 22959341]
4. Doody MM, Hayes HM, Bilgrad R. Comparability of national death index plus and standard procedures for determining causes of death in epidemiologic studies. Ann Epidemiol 2001 Jan;11(1):46-50. [Medline: 11164119]
5. Wentworth DN, Neaton JD, Rasmussen WL. An evaluation of the Social Security Administration master beneficiary record file and the National Death Index in the ascertainment of vital status. Am J Public Health 1983 Nov;73(11):1270-1274. [Medline: 6625030]

6. Dredger SM, Kothari A, Morrison J, Sawada M, Crighton EJ, Graham ID. Using participatory design to develop (public) health decision support systems through GIS. Int J Health Geogr 2007 Nov 27;6:53 [FREE Full text] [doi: 10.1186/1476-072X-6-53] [Medline: 18042298]
7. Gao S, Mioc D, Anton F, Yi X, Coleman DJ. Online GIS services for mapping and sharing disease information. Int J Health Geogr 2008 Feb 25;7:8 [FREE Full text] [doi: 10.1186/1476-072X-7-8] [Medline: 18298859]
8. van der Wilk EA, Verschuuren M. Adding value to figures: a web-based European public health information system. J Telemed Telecare 2010;16(1):35-41. [doi: 10.1258/jtt.2009.001010] [Medline: 20086266]
9. Koenig A, Samarasundera E, Cheng T. Interactive map communication: pilot study of the visual perceptions and preferences of public health practitioners. Public Health 2011 Aug;125(8):554-560. [doi: 10.1016/j.puhe.2011.02.011] [Medline: 21782202]
10. Ballot Pedia. 2016. Missouri State Senate URL: https://ballotpedia.org/Missouri_State_Senate [accessed 2017-01-04] [WebCite Cache ID 6nGokdD7M]
11. Missouri Senate. 2016. Missouri Senate-Statewide District Map URL: http://www.senate.mo.gov/districtmap/ [accessed 2017-04-05] [WebCite Cache ID 6pVIDM0fj]
12. University of Missouri Health School of Medicine. 2017. Missouri Cancer Registry and Research Center URL: http://mcr.umh.edu/ [accessed 2017-08-22] [WebCite Cache ID 6suUbiBxu]
13. Centers for Disease Control Prevention. National Center for Health Statistics: National Death Index; 2017. URL: https://www.cdc.gov/nchs/ndi/index.htm [WebCite Cache ID 6r2oZLWEF]
14. Esri. 2016. ArcMap URL: http://desktop.arcgis.com/en/arcmap/ [WebCite Cache ID 6nGrs5PKj]
15. United States Census Bureau. 2016. TIGER/Line Shapefiles and TIGER/Line Files URL: https://www.census.gov/geo/maps-data/data/tiger-line.html [accessed 2017-04-05] [WebCite Cache ID 6pVITErPP]
16. All About Redistricting: Professor Justin Levitt's Guide to Drawing the Electoral Lines. 2017. Missouri URL: http://redistricting.lls.edu/states-MO.php [accessed 2017-06-05] [WebCite Cache ID 6qzlylmOb]
17. National Cancer Institute: Surveillance, Epidemiology, and End Results Program. SEER*Stat Software URL: https://seer.cancer.gov/seerstat/ [accessed 2017-01-04] [WebCite Cache ID 6nGp6lg5B]
18. National Cancer Institute: Surveillance, Epidemiology, and End Results Program. 2016. Cancer statistics URL: https://seer.cancer.gov/causespecific/ [accessed 2017-04-05] [WebCite Cache ID 6pVIYnVI3]
19. Hofferkamp J. Standards for Cancer Registries Volume III: Standards for Completeness, Quality, Analysis, Management, Security and Confidentiality of Data. Springfield, IL: North American Association of Central Cancer Registries; 2008 Aug. URL: https://20tqtx36s1la18rvn82wcmpn-wpengine.netdna-ssl.com/wp-content/uploads/2016/11/Standards-for-Completeness-Quality-Analysis-Management-Security-and-Confidentiality-of-Data-August-2008PDF.pdf [WebCite Cache ID 6r2pEprbo]
20. United States Census Bureau. 2017. Cartographic boundary shapefiles URL: https://www.census.gov/geo/maps-data/data/tiger-cart-boundary.html [WebCite Cache ID 6pVIfq99X]
21. ESRI Ltd. 2016. Interactive mapping and reporting services for data professionals URL: http://www.instantatlas.com/ [WebCite Cache ID 6nGrPHQ6f]
22. North America Association of Central Cancer Registries. Chapter X: Data Dictionary, Version 16. URL: http://datadictionary.naaccr.org/default.aspx?c=10#3020 [WebCite Cache ID 6sjzLV0PR]
23. MCR-ARC InstantAtlas Cancer Data Portal. 2016. MCR-ARC County and senate district by age, race, stage, & grade area health profile report URL: https://instantatlas.umh.edu/IAS/DynamicReports/areaHealthProfile2015-SD/atlas.html [WebCite Cache ID 6pVIoIvwd]
24. MCR-ARC InstantAtlas Cancer Data Portal. 2016. MCR-ARC County and senate district by age, race, stage, & grade double map report URL: https://instantatlas.umh.edu/IAS/DynamicReports/doubleMap2015-SD/atlas.html [WebCite Cache ID 6pVJglkrE]
25. Ellis L, Woods LM, Estève J, Eloranta S, Coleman MP, Rachet B. Cancer incidence, survival and mortality: explaining the concepts. Int J Cancer 2014 Oct 15;135(8):1774-1782 [FREE Full text] [doi: 10.1002/ijc.28990] [Medline: 24945976]
26. Curado M, Edwards B, Shin H, Storm H, Ferlay J, Heanue M, et al. Cancer Incidence in Five Continents, Volume IX. Lyon, France: IARC Press, International Agency for Research on Cancer; 2007. URL: https://www.iarc.fr/en/publications/pdfs-online/epi/sp160/CI5vol9-A.pdf [WebCite Cache ID 6nGssYU3M]
27. Restore@National Centre for Research Methods. 2017. Geographical referencing learning resources: the modifiable areal unit problem URL: http://www.restore.ac.uk/geo-refer/91023cwors00y00000000.php [WebCite Cache ID 6usjBe6u0]
28. Missouri Department of Health & Senior Services. 2015. The Behavioral Risk Factor Surveillance System (BRFSS) URL: http://health.mo.gov/data/brfss/index.php [WebCite Cache ID 6nGt6DCjO]
29. Missouri House of Representatives. 2017. Missouri redistricting maps adopted 2012: statewide House district map URL: http://www.house.mo.gov/billtracking/maps/house.pdf [WebCite Cache ID 6pVJZTKZu]
30. Ballot Pedia. 2017. Missouri House of Representatives District 163 URL: https://ballotpedia.org/Missouri_House_of_Representatives_District_163 [accessed 2017-05-09] [WebCite Cache ID 6qKkcxatI]

31. Ben Ramadan AA, Jackson-Thompson J, Schmaltz CL. Usability assessment of the Missouri Cancer Registry's published interactive mapping reports: round one. JMIR Hum Factors 2017 Aug 04;4(3):e19 [FREE Full text] [doi: 10.2196/humanfactors.7899] [Medline: 28778842]

Abbreviations

NDI: National Death Index
SEER: Surveillance, Epidemiology, and End Results
SSDI: Social Security Death Index

Edited by G Eysenbach; submitted 07.06.17; peer-reviewed by C Johnson, E Nelson; comments to author 28.06.17; revised version received 22.08.17; accepted 17.01.18; published 03.05.18.

Please cite as:
Ben Ramadan AA, Jackson-Thompson J, Schmaltz CL
Improving Visualization of Female Breast Cancer Survival Estimates: Analysis Using Interactive Mapping Reports
JMIR Public Health Surveill 2018;4(2):e42
URL: http://publichealth.jmir.org/2018/2/e42/
doi:10.2196/publichealth.8163
PMID:29724710

Original Paper

Economic Recession and Obesity-Related Internet Search Behavior in Taiwan: Analysis of Google Trends Data

Ho-Wei Wang[1*], PhD, MBA; Duan-Rung Chen[2*], PhD

[1]Institute of Health Policy and Management, National Taiwan University, Taipei, Taiwan
[2]Institute of Health Behaviors and Community Sciences, National Taiwan University, Taipei, Taiwan
[*]all authors contributed equally

Corresponding Author:
Duan-Rung Chen, PhD
Institute of Health Behaviors and Community Sciences
National Taiwan University
6F, No 17, Xu-Zhou Road
Taipei, 100
Taiwan
Phone: 886 2 33668066
Fax: 886 2 33668063
Email: duan@ntu.edu.tw

Abstract

Background: Obesity is highly correlated with the development of chronic diseases and has become a critical public health issue that must be countered by aggressive action. This study determined whether data from Google Trends could provide insight into trends in obesity-related search behaviors in Taiwan.

Objective: Using Google Trends, we examined how changes in economic conditions—using business cycle indicators as a proxy—were associated with people's internet search behaviors related to obesity awareness, health behaviors, and fast food restaurants.

Methods: Monthly business cycle indicators were obtained from the Taiwan National Development Council. Weekly Taiwan Stock Exchange (TWSE) weighted index data were accessed and downloaded from Yahoo Finance. The weekly relative search volumes (RSV) of obesity-related terms were downloaded from Google Trends. RSVs of obesity-related terms and the TWSE from January 2007 to December 2011 (60 months) were analyzed using correlation analysis.

Results: During an economic recession, the RSV of obesity awareness and health behaviors declined (r=.441, P<.001; r=.593, P<.001, respectively); however, the RSV for fast food restaurants increased (r=−.437, P<.001). Findings indicated that when the economy was faltering, people tended to be less likely to search for information related to health behaviors and obesity awareness; moreover, they were more likely to search for fast food restaurants.

Conclusions: Macroeconomic conditions can have an impact on people's health-related internet searches.

(JMIR Public Health Surveill 2018;4(2):e37) doi:10.2196/publichealth.7314

KEYWORDS
obesity; economic recession; Google Trends; fast food; internet search; health-seeking behaviors; infodemiology

Introduction

Background

Obesity is highly correlated with the development of several chronic diseases, including cardiovascular disease and type 2 diabetes [1,2]. According to the World Obesity Federation, if levels of obesity and overweight status continue to rise, 2.7 billion adults will be overweight by 2025, up from 2 billion in 2014 [3]. In Taiwan, the prevalence of individuals defined as overweight and obese is also increasing with an estimated prevalence of 48.9% in males and 38.3% in females [4]. Taking the obesity rates of both adults and children into account, Taiwan ranks highest in Asia, suggesting a critical public health issue that must be solved decisively [4].

Previous Research

Studies have found that the prevalence of obesity shifts toward lower socioeconomic status groups as a country's gross national product increases [5,6]. Studies conducted in the United States and Europe have determined that the highest obesity rates are associated with the lowest income levels [7,8]. One reason for increased obesity among those with lower incomes is that a lower income is often associated with poor health behavior and dietary quality [9,10].

Several studies have also indicated a potential association between macroeconomic conditions and obesity-related health behavior. For example, one study used a convenience sample from a metropolitan city in the US Midwest and discovered that an economic downturn may increase rates of obesity by forcing families to cut their food expenditure [11,12]. Moreover, the downturn motivated consumers to switch to lower-priced and higher-calorie foods, which usually have higher fat and sugar content while containing lower quantities of recommended micronutrients [11,12]. Another study in the United States found that recession-driven unemployment was associated with an increase in higher-calorie purchases [13]. Other studies, using the US Behavioral Risk Factor Surveillance System data, revealed that a recession can lead to a lower dietary quality and less physical activity [14,15]. However, other studies have indicated that people have more time to exercise and prepare healthy meals during economic downturns and are thus more likely to maintain a healthy weight [16,17].

The most recent global recession (Great Recession, 2008-2009) was the most severe, as measured by duration, since the Great Depression of the 1930s [18]. In Taiwan, several indices reached record lows during the most recent recession. Car sales declined 28.09% in September 2008, retail sales fell to a 7-year low in September 2008 (–5.06% year-on-year growth), the unemployment rate soared to a 4-year high (4.27%), and the overall consumer confidence index dropped to 79.56 [19,20]. All of these figures revealed that consumer spending had been retrenched by the decline in income [19,20]. Although various sources of conflicting evidence regarding the effect of the recession on health behaviors remain controversial, recessions still provide a unique opportunity to shed light on the association between economic conditions and the risk of obesity.

Individuals today often use the internet and various search engines, such as Google and Yahoo, to obtain information to support their decisions [21]. These online searches through search engines create trend data, which can be analyzed in real time [22,23]. Since 2004, Google has provided two services for trend analyses: Google Flu Trends and Google Trends. Google Flu Trends has been identified as a powerful tool used in influenza surveillance in the United States, identifying influenza epidemics up to 7-10 days before detection by the Centers for Disease Control and Prevention's influenza surveillance network [22]. During periods of infectious disease prevalence, Google Flu Trends has also been able to predict the volume of emergency department visits [24,25]. However, Google Flu Trends failed to correctly estimate the scale of the 2009 H1N1 pandemic in the United States. Several mistakes that led to Google Flu Trends' overestimation of H1N1 incidence were caused by the limited transparency of Google's treatment of data and its dynamic algorithm, which was due to Google's business considerations [26,27]. The Google Flu Trends service is no longer available in many countries.

Google Trends also has predictive capacity for monitoring the epidemic curves of food-borne illnesses, such as peanut butter-associated outbreaks of *Salmonella enterica* serotype Typhimurium [28] and the incidence of human immunodeficiency virus [29,30]. With regard to population health concerns during the Great Recession, a study using Google Trends found that specific searches with the greatest relative excess were stomach ulcer and headache symptoms [31].

Research Goals

The purpose of this study was not to contest or generate additional theories to support the contentious link between macroeconomic conditions and obesity-related health behaviors, but rather it was to understand how an economic downturn might affect obesity-related internet search behavior. Thus, by using Google Trends we investigated how changes in economic conditions—using business cycle indicators as a proxy—affected people's internet search behaviors, including searches related to obesity awareness, obesity-related health behavior, and fast food restaurants.

Methods

Data Sources

Great Recession Period

The 2008-2009 economic recession was the most severe since the Great Depression of the 1930s; therefore, it provides an excellent opportunity to study how the business cycle affects internet search behaviors for health-related issues [18]. Business cycle indicators, released monthly by the Taiwan National Development Council, are used to measure economic development [32]. The scores of indicators represent five different levels of economic prosperity, including sluggish (scores of 9-16), downward transition (scores of 17-22), stable (scores of 23-31), upward transition (scores of 32-37), and booming (scores of 38-45) [32] (see Figure 1 for further details).

In this study, the recession period was defined as the time when the business cycle indicator dropped below the stable stage, climbed to the transitional stage (above 23) from April 2008 through September 2009, and then returned to growth until December 2011 (see Figure 2 for graphical data).

Google Trends Search Terms and Trends

Google Trends analyzes all search queries of a specific term to quantify interest in topics at the population level, thus serving as an increasingly useful research tool. Relative search volume (RSV) is the proportion of searches for a given term out of all searches for a given geographic location and time period, which is then normalized to a 0-100 scale [33,34]. To capture a broad sense of the search terms regarding obesity-related conditions and behaviors, 49 search terms were cataloged through Google Trends; the relevant terms were derived from Google Trends' explore function [31]. The health belief model (HBM) is a psychological health-behavior-change model developed to explain and predict health-related behaviors, in particular the uptake of health services [35]. HBM has been applied to understand patients' responses to symptoms of disease, compliance with medical regimens, and lifestyle behaviors. In this study, we classified these key search terms into three categories: (1) obesity-related awareness terms, such as metabolic syndrome; (2) obesity-related health behavior terms, such as healthy diet; and (3) fast food restaurant terms such as McDonald’s (Table 1).

Google Trends only reports results above a certain threshold. When Google Trends cannot report the search volume for a term, the message displayed is, “Not enough search volume to show graphs.” Thirty-four search terms had insufficient search volume. Weekly data on the remaining 15 search terms from January 2007 to December 2011 (60 months) were accessed and downloaded from Google Trends on June 28, 2016. Table 1 displays the list of obesity-related search terms categorized into three groups. Multimedia Appendix 1 lists these terms in Chinese, along with their English equivalents.

Figure 1. Taiwan business cycle indicators.

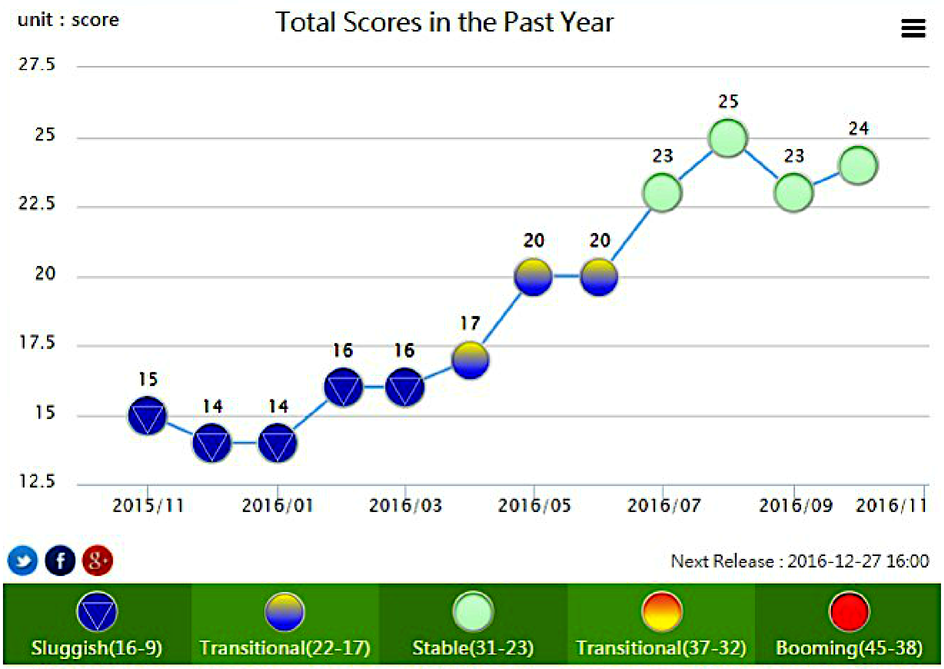


Figure 2. Time series for the monthly business cycle indicator and Taiwan Stock Exchange Weighted Index (TWSE) between January 2006 and December 2011.

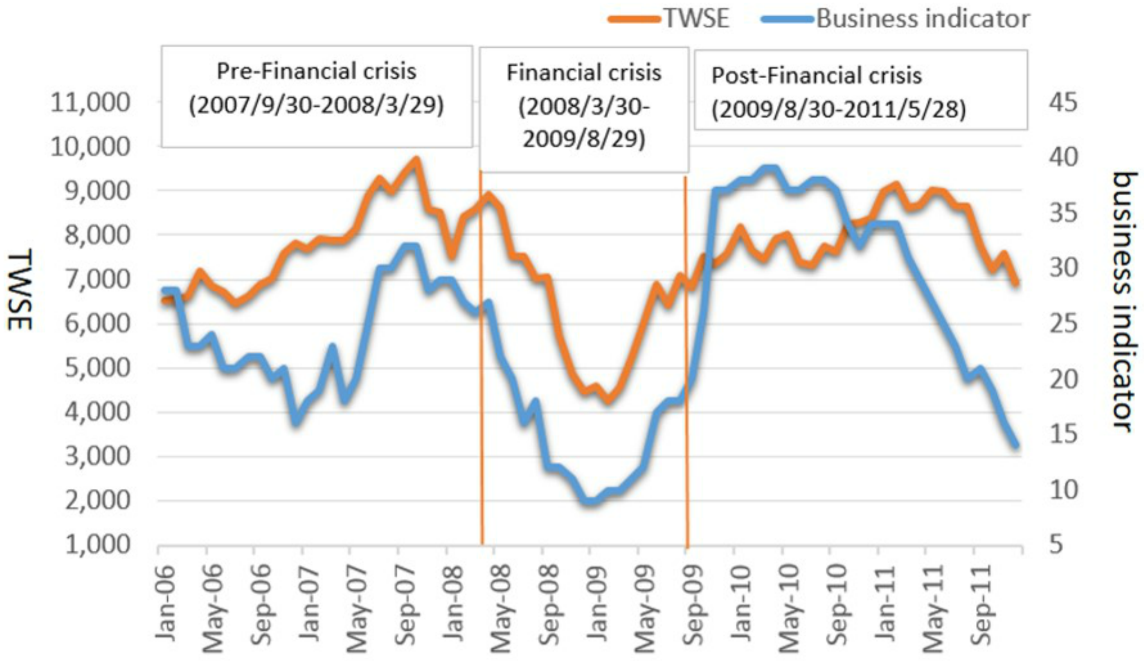


Table 1. List of obesity-related search terms.

Category	Search terms
Awareness terms	Obesity; metabolic syndrome; endocrinology; triglycerides; body fat percentage; body mass index
Health behavior terms	Healthy diet; aerobics; effective diet; meal replacement; on a diet; fitness
Fast food restaurant terms	McDonald's; KFC; pizza

Taiwan Stock Exchange Weighted Index

As stock market volatility is a predictor for macroeconomic volatility [36,37], we demonstrated the relationship between the Taiwan Stock Exchange (TWSE) and the RSVs of obesity-related terms during three business cycles. Weekly TWSE data were accessed and downloaded from the Taiwan Stock Exchange Corporation on June 30, 2016 [38].

Statistical Analyses

Pearson correlation coefficient analysis is widely used in Google Trends research [39]. To assess the strength of the linear relationship between the RSVs and TWSE, a Pearson correlation coefficient analysis with 95% CI was reported. The correlation between the RSVs and TWSE was analyzed first and then followed by lead pattern analysis. The TWSE's lead pattern analysis was then analyzed by 1-month and 3-month intervals. For example, a 1-month lead evaluated the correlation of the TWSE from January 2009 with obesity-related searches in February 2009. All statistical analyses were conducted using IBM SPSS Statistics for Windows, version 20.0 (IBM Corp, Armonk, NY); a two-tailed *P* value of less than .05 was required for statistical significance in all analyses conducted.

Results

Table 2 lists the change of RSVs by categories. The RSV of obesity awareness terms showed a continual downward trend across the three periods. Despite the upturn in the financial market, the RSV of obesity awareness terms became weaker during the postfinancial crisis period, and the RSV of health behavior terms fell dramatically (by 12%) during the financial crisis period. However, the RSV of fast food restaurants increased sharply during this period, rising by 27% (see Figure 3). Autoregressive integrated moving average (ARIMA) models were applied to predict the tendency by the end of April 2010. The results reported that awareness and behavior terms trended down, while fast food terms were showed an upward trend (see Figure 3).

Figure 3 shows time trends before and during the Great Recession for all three categories of search terms that were analyzed. Each line indicates a specific query trend, with the red line indicating the tendency across all queries based on ARIMA models.

Nowcasting Effect

Nowcasting, also known as contemporaneous forecasting, is used to describe the extent of a current association. During the prefinancial crisis period, the RSV of health behaviors was positively associated with the TWSE (r=.504, P=.009), suggesting that health behavior searches increased while the TWSE went up. The RSV of fast food restaurants was negatively associated with the TWSE (r=−.529, P=.006), indicating that fast food restaurant searches increased while the TWSE declined.

Second, during the period of the financial crisis, all RSVs of the three categories were correlated with the TWSE. The RSV of obesity awareness and health behavior declined while the TWSE declined (r=.441, P<.001; r=.593, P<.001, respectively), and fast food restaurant searches increased while the TWSE declined (r=−.437, P<.001).

Third, the RSV of obesity awareness and health behavior correlated with the TWSE after the financial crisis, but the correlation coefficient changed from a positive to a negative direction (r=−.544, P<.001; r=−.548, P<.001). This finding may suggest that people were less concerned with obesity-related issues when job opportunities returned during the postfinancial crisis period (see Table 3).

Forecasting Effect

In addition to the nowcasting effect, the forecasting effect can be used to determine the TWSE's lead pattern over the search trends. As aforementioned, for example, a 1-month lead TWSE represented the correlation between the TWSE in January 2009 and obesity-related searches in February 2009. All figures are reported in Table 3.

Table 2. Relative search volumes (RSVs) by categories.

Time period	Search volume indexes (RSVs)		
	Awareness (% change)	Behavior (% change)	Fast food (% change)
Reference (before prefinancial crisis; 2006/10/1-2007/9/29)	34	23	23
Prefinancial crisis (2007/9/30-2008/3/29)	32 (–6)	23 (0)	27 (16)
Financial crisis (2008/3/30-2009/8/29)	30 (–6)	20 (–12)	34 (27)
Postfinancial crisis (2009/8/30-2011/5/28)	29 (–3)	19 (–5)	38 (12)

Figure 3. Time trends for queries of the three categories around the Great Recession.

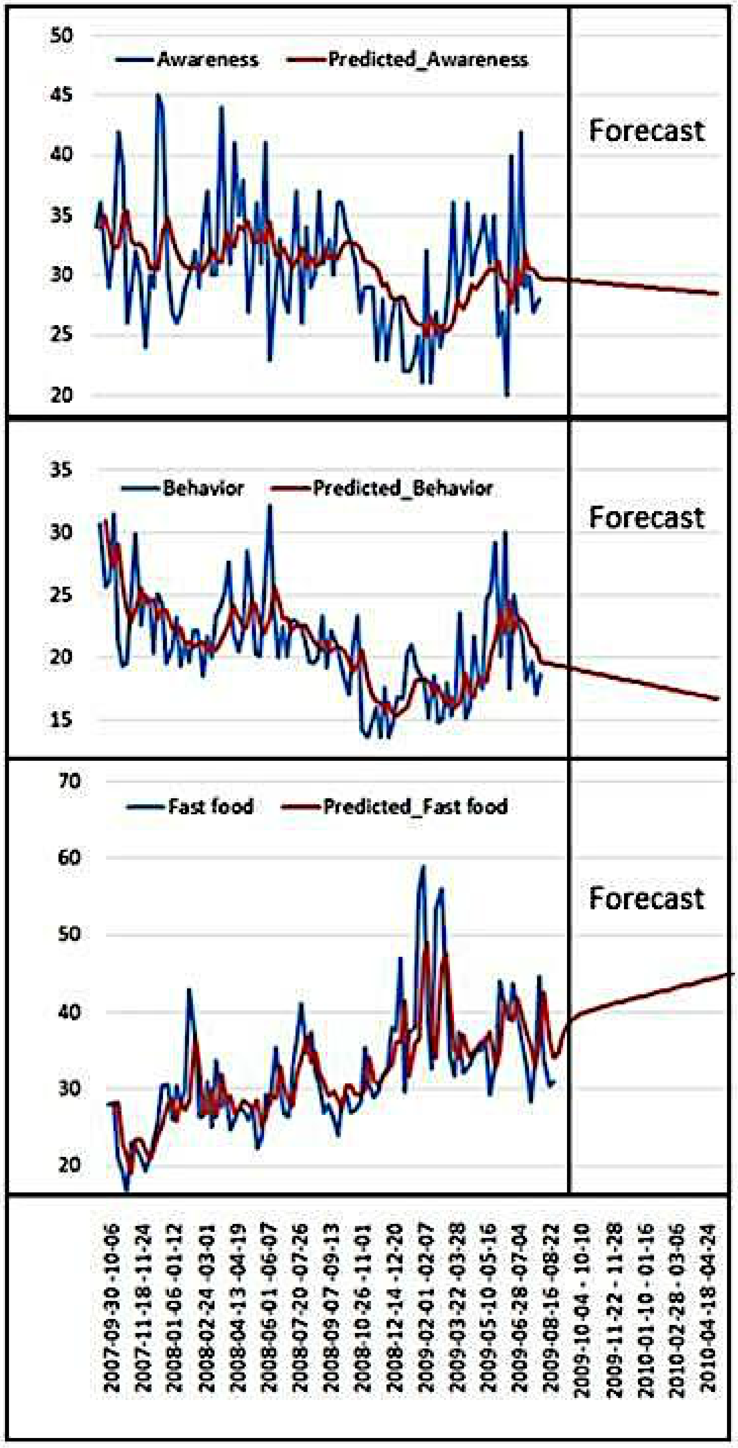


Table 3. Pearson cross-correlation analysis of search terms and Taiwan Stock Exchange (TWSE), 2007-2011.

Attribute and time period	TWSE preceded search terms by time (*P* value)		
	No lag	1 month	3 months
Obesity awareness			
Prefinancial crisis (2007/9/30-2008/3/29)	.277 (.170)	.086 (.677)	.087 (.674)
Financial crisis (2008/3/30-2009/8/29)	.441 (<.001)	.453 (<.001)	.318 (.006)
Postfinancial crisis (2009/8/30-2011/5/28)	–.554 (<.001)	–.595 (<.001)	–.354 (.001)
Health Behaviors			
Prefinancial crisis (2007/9/30-2008/3/29)	.504 (.009)	.421 (.032)	.344 (.086)
Financial crisis (2008/3/30-2009/8/29)	.593 (<.001)	.646 (<.001)	.388 (.001)
Postfinancial crisis (2009/8/30-2011/5/28)	–.548 (<.001)	–.405 (<.001)	–.298 (.004)
Fast Food			
Prefinancial crisis (2007/9/30-2008/3/29)	–.529 (.006)	–.640 (<.001)	–.028 (.893)
Financial crisis (2008/3/30-2009/8/29)	–.437 (<.001)	–.513 (<.001)	–.530 (<.001)
Postfinancial crisis (2009/8/30-2011/5/28)	.049 (.646)	.095 (.372)	.181 (.086)

First, during the prefinancial crisis period, a 1-month lead TWSE was statistically associated with the RSVs of health behaviors and fast food (*r*=.421, *P*=.032; *r*=−.640, *P*<.001, respectively). Second, during the financial crisis period, a 3-month lead TWSE was found to be significantly associated with the RSVs of all three categories (obesity awareness: *r*=.318, *P*=.006; health behaviors: *r*=.388, *P*=.001; fast food: *r*=−.530, *P*<.001).

For searches using obesity awareness and health behavior terms, the TWSE showed a 3-month predictive power with positive association, suggesting that the stock market collapse was followed 3 months later by a decline in interest for obesity awareness and health behaviors. These results imply that people were more likely to disregard obesity-related health issues when the business cycle fell into recession. Furthermore, the TWSE was negatively related to the RSV of fast food restaurants (*r*=−.530, *P*<.001), suggesting that searches for fast food were possibly caused by a sluggish economy and consequently limited budgets for healthy food.

Third, after the financial crisis period, although a business cycle and stock market turnaround took place, searches for obesity awareness and health behavior did not follow the same trend. In contrast, both categories of search terms demonstrated a negative correlation with TWSE, suggesting that obesity-related health issues may have been lower than before the economic downturn due to the economic recovery, which lead to job opportunity increases (lower unemployment).

Finally, fast food consumption has been assumed to be a potential risk factor for being overweight and obese, due to the high energy densities and high glycemic loads associated with such foods [40]. Research has suggested that an economic downturn pushes individuals to replace nutritious but expensive foods with cheaper and higher-calorie substitutes [15,41]. Such findings can partially be explained by lower budgets for food expenditure [11,42]. Moreover, commercial promotions by fast food companies during the recession periods might have triggered the consumption of fast foods; a trend that was indicated by the stock price of fast food companies (eg, McDonald’s) that bucked the negative trend while the major index continued to struggle (see Figure 4) [43,44].

Our findings support this relationship by demonstrating the negative correlation between TWSE and fast food searches during the financial crisis period (*r*=−.437, *P*<.001).

The consumer price index (CPI) tracks monthly data on changes in the prices paid by consumers for a specific or representative basket of goods and services [45]. We analyzed the relationship between unemployment and CPI for meat and fruit in Taiwan from 2008 through 2016 (see Figure 5); the results revealed a positive association between unemployment rate and meat consumption (*r*=.209, *P*=.032) and negative association between unemployment rate and fruit consumption (*r*=−.421, *P*<.001; see Table 4). Such a result echoes studies that indicated people increase caloric purchases, such as meat, for protein satiety during an economic downturn (high unemployment rate) and spend less on healthy foods, such as fruit [11,13,41,42]. Furthermore, the CPI for cigarettes and betel nuts (a commonly chewed nut in Taiwan with a mildly stimulating effect) also had a positive relationship with unemployment (*r*=.220, *P*=.024), which suggests unhealthy behavior may also rise when the economy and job market are sluggish.

Figure 4. S&P 500 and McDonald’s Corp. (MCD:NYSE) share performance between 2007 and 2010.

Figure 5. Consumer price index and unemployment trends, 2008-2016.

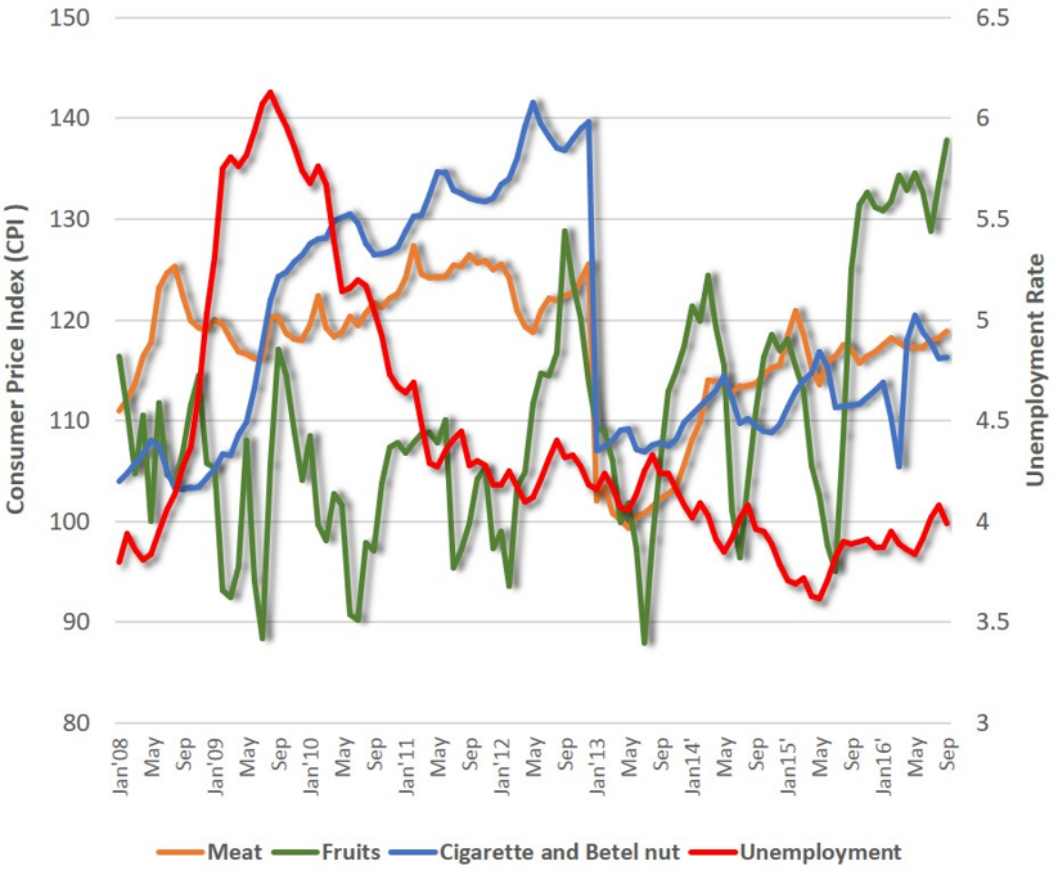


Table 4. Pearson cross-correlation analysis of the correlation between consumer price index and unemployment for meat, fruit, and cigarettes and betel nuts (2008-2016).

Variable	Unemployment	Meat	Fruit	Cigarette and betel nuts
Unemployment	1.00			
Meat (*P* value)	.209 (0.32)	1.00		
Fruit (*P* value)	–.421 (<.001)	.003 (.974)	1.00	
Cigarettes and betel nuts (*P* value)	.220 (.024)	.599 (<.001)	–.094 (.339)	1.00

Discussion

Principal Findings

This study contributes to an understanding of how business cycles are related to online searches for obesity-related awareness, health behavior, and fast food restaurants. The results revealed a positive association between the TWSE, obesity-related awareness, and health behavior searches during the prefinancial and financial crisis periods, suggesting that economic hardship may be the underlying force affecting individual obesity-related internet searches. Previous studies have indicated that individuals may engage in less physical activity or cease membership of health and sports clubs during periods of economic downturn [14,15,46]. This study revealed that during the postfinancial crisis period, the associations between the TWSE and searches for obesity awareness and obesity-related health behaviors changed from positive to negative, suggesting that as the TWSE increased, internet searches for obesity awareness and obesity-related health behaviors decreased. Such a finding might be explained by job stress or having less time for self-care activities during economic upturns [17,47]. However, despite the upturn in the financial market, the RSV of obesity awareness terms became weaker during the postfinancial crisis period (see Table 2). This finding may suggest that during the postfinancial crisis period, people had more job opportunities and less time to be concerned with obesity-related issues.

The results suggested that the performance of the TWSE moderately correlated with these search trends, demonstrating the possibility that monitoring the TWSE could plausibly predict future obesity-related searches on the internet. Additionally, the results also suggested that the TWSE demonstrated a 3-month leading effect; however, the 1-month leading effect was much stronger during the prefinancial crisis and financial crisis periods.

Finally, our results suggested that the TWSE could also play a role in monitoring changes in obesity-related aggregated individual search behaviors at population scale. If so, one of the primary indicators of economic prosperity not only represents the nation’s economy but is also a leading indicator of population health. This study provides a new methodological lens for improving the monitoring of obesity. This method can help governments recognize that recessions may result in an increase in obesity-related problems. While addressing the issues of a stalling job market and high unemployment, governments should also look to provide resources for dealing with obesity.

Limitations

Due to the Google Trends algorithm, several limitations of this study should be noted. First, there was difficulty in identifying search trends that were generated by true cases. In particular, Google Trends tends to be influenced by media exposure of specific diseases (eg, drug advertisements), which drives more nonrelated individuals to search for terms and thus increases the search volume [48]. This limitation can possibly be more effectively controlled in future research [49]. Another limitation was that the calculation of Google Trends depends on Google’s assumptions and normalization, which are not clearly reported. Third, the method by which relevant terms are derived from Google Trends’ explore function has also not been disclosed by Google. Finally, potential alternative variables, such as increased advertising and low cost special offers by fast food companies during economic recessions may be confounders that influenced the results; this issue is worthy of study in future research.

Conclusion

In conclusion, internet search data can be a potentially useful tool for health policy makers to identify obesity-related issues and possible obesity problems within the population. Google Trends serves as an easily accessible and real-time surveillance tool. Despite its limitations, this study highlighted Google Trends as a useful tool for establishing a plausible relationship between obesity-related search terms and macroeconomic events.

Conflicts of Interest

None declared.

Multimedia Appendix 1

List of Chinese obesity-related search terms and their English equivalents.

[PNG File, 51KB - publichealth_v4i2e37_app1.png]

References

1. Must A, Spadano J, Coakley EH, Field AE, Colditz G, Dietz WH. The disease burden associated with overweight and obesity. JAMA 1999 Oct 27;282(16):1523-1529. [Medline: 10546691]
2. Burton BT, Foster WR, Hirsch J, Van ITB. Health implications of obesity: an NIH Consensus Development Conference. Int J Obes 1985;9(3):155-170. [Medline: 3840463]
3. World Obesity Federation. About Obesity. 2015. URL: http://www.worldobesity.org/site_media/uploads/World_Obesity_Day_Press_Release.pdf [accessed 2017-01-11] [WebCite Cache ID 6nPdaocDP]
4. Health Promotion Administration. Ministry of Health and Welfare. 2016 Dec 28. 2013-2014 Nutrition and Health Survey in Taiwan (NAHSIT) URL: http://www.hpa.gov.tw/English/index.aspx [accessed 2016-06-20] [WebCite Cache ID 6nPdp5Jid]
5. Monteiro CA, Moura EC, Conde WL, Popkin BM. Socioeconomic status and obesity in adult populations of developing countries: a review. Bull World Health Organ 2004 Dec;82(12):940-946 [FREE Full text] [Medline: 15654409]
6. Chen D, Truong K. Using multilevel modeling and geographically weighted regression to identify spatial variations in the relationship between place-level disadvantages and obesity in Taiwan. Applied Geography 2012;32(2):745. [doi: 10.1016/j.apgeog.2011.07.018]
7. Schoenborn CA, Adams PF, Barnes PM. Body weight status of adults: United States, 1997-98. Adv Data 2002 Sep 06(330):1-15. [Medline: 12661975]
8. Lennernäs M, Fjellström C, Becker W, Giachetti I, Schmitt A, Remaut DWA, et al. Influences on food choice perceived to be important by nationally-representative samples of adults in the European Union. Eur J Clin Nutr 1997 Jun;51 Suppl 2:S8-15. [Medline: 9222718]
9. Darmon N, Ferguson EL, Briend A. A cost constraint alone has adverse effects on food selection and nutrient density: an analysis of human diets by linear programming. J Nutr 2002 Dec;132(12):3764-3771 [FREE Full text] [Medline: 12468621]
10. Irala-Estévez JD, Groth M, Johansson L, Oltersdorf U, Prättälä R, Martínez-González MA. A systematic review of socio-economic differences in food habits in Europe: consumption of fruit and vegetables. Eur J Clin Nutr 2000 Sep;54(9):706-714. [Medline: 11002383]
11. Organisation for Economic Co-operation and Development. Obesity Update 2014. 2014 Jun 20. URL: http://www.oecd.org/health/Obesity-Update-2014.pdf [accessed 2018-03-28] [WebCite Cache ID 6yGbYIOLv]
12. Miller CK, Branscum P. The effect of a recessionary economy on food choice: implications for nutrition education. J Nutr Educ Behav 2012;44(2):100-106. [doi: 10.1016/j.jneb.2011.01.015] [Medline: 21900050]
13. Ng SW, Slining MM, Popkin BM. Turning point for US diets? Recessionary effects or behavioral shifts in foods purchased and consumed. Am J Clin Nutr 2014 Mar;99(3):609-616 [FREE Full text] [doi: 10.3945/ajcn.113.072892] [Medline: 24429538]
14. Zhang Q, Lamichhane R, Wang Y. Associations between U.S. adult obesity and state and county economic conditions in the recession. J Clin Med 2014 Jan 27;3(1):153-166 [FREE Full text] [doi: 10.3390/jcm3010153] [Medline: 26237254]
15. Ludwig DS, Pollack HA. Obesity and the economy: from crisis to opportunity. JAMA 2009 Feb 04;301(5):533-535. [doi: 10.1001/jama.2009.52] [Medline: 19190318]
16. Ruhm CJ. Healthy living in hard times. J Health Econ 2005 Mar;24(2):341-363. [doi: 10.1016/j.jhealeco.2004.09.007] [Medline: 15721049]
17. Ruhm CJ. Are Recessions Good for Your Health? Q J Econ 2000 May 01;115(2):617-650. [doi: 10.1162/003355300554872]
18. National Bureau of Economic Research. U.S. business cycle expansions and contractions. 2016. URL: http://www.nber.org/cycles.html [accessed 2016-06-18] [WebCite Cache ID 5QOwh5ffT]
19. National Policy Foundation. The impact of financial crisis on Taiwan economic development and strategy. 2000. URL: http://www.npf.org.tw/3/4985 [accessed 2016-06-15] [WebCite Cache ID 6nPfU7kkp]
20. Automotive Research and Testing Center. Monthly Auto Sales Report. 2010. URL: https://www.artc.org.tw/chinese/03_service/03_02detail.aspx?pid=1455 [accessed 2016-06-15] [WebCite Cache ID 6nPfaR21g]
21. Moat HS, Preis T, Olivola CY, Liu C, Chater N. Using big data to predict collective behavior in the real world. Behav Brain Sci 2014 Feb;37(1):92-93. [doi: 10.1017/S0140525X13001817] [Medline: 24572233]
22. Ginsberg J, Mohebbi MH, Patel RS, Brammer L, Smolinski MS, Brilliant L. Detecting influenza epidemics using search engine query data. Nature 2009 Feb 19;457(7232):1012-1014. [doi: 10.1038/nature07634] [Medline: 19020500]
23. Eysenbach G. Infodemiology and infoveillance: framework for an emerging set of public health informatics methods to analyze search, communication and publication behavior on the Internet. J Med Internet Res 2009;11(1):e11 [FREE Full text] [doi: 10.2196/jmir.1157] [Medline: 19329408]
24. Malik M, Gumel A, Thompson L, Strome T, Mahmud S. "Google flu trends" and emergency department triage data predicted the 2009 pandemic H1N1 waves in Manitoba. Can J Public Health 2011 Aug 01;102(4):294-297. [Medline: 21913587]
25. Dugas AF, Hsieh Y, Levin SR, Pines JM, Mareiniss DP, Mohareb A, et al. Google Flu Trends: correlation with emergency department influenza rates and crowding metrics. Clin Infect Dis 2012 Feb 15;54(4):463-469 [FREE Full text] [doi: 10.1093/cid/cir883] [Medline: 22230244]
26. Lazer D, Kennedy R, King G, Vespignani A. Big data. The parable of Google Flu: traps in big data analysis. Science 2014 Mar 14;343(6176):1203-1205. [doi: 10.1126/science.1248506] [Medline: 24626916]

27. Cook S, Conrad C, Fowlkes AL, Mohebbi MH. Assessing Google flu trends performance in the United States during the 2009 influenza virus A (H1N1) pandemic. PLoS One 2011;6(8):e23610 [FREE Full text] [doi: 10.1371/journal.pone.0023610] [Medline: 21886802]
28. Brownstein JS, Freifeld CC, Madoff LC. Digital disease detection--harnessing the Web for public health surveillance. N Engl J Med 2009 May 21;360(21):2153-5, 2157 [FREE Full text] [doi: 10.1056/NEJMp0900702] [Medline: 19423867]
29. Jena AB, Karaca-Mandic P, Weaver L, Seabury SA. Predicting new diagnoses of HIV infection using Internet search engine data. Clin Infect Dis 2013 May;56(9):1352-1353 [FREE Full text] [doi: 10.1093/cid/cit022] [Medline: 23334812]
30. Zheluk A, Quinn C, Hercz D, Gillespie JA. Internet search patterns of human immunodeficiency virus and the digital divide in the Russian Federation: infoveillance study. J Med Internet Res 2013;15(11):e256 [FREE Full text] [doi: 10.2196/jmir.2936] [Medline: 24220250]
31. Althouse BM, Allem J, Childers MA, Dredze M, Ayers JW. Population health concerns during the United States' Great Recession. Am J Prev Med 2014 Feb;46(2):166-170. [doi: 10.1016/j.amepre.2013.10.008] [Medline: 24439350]
32. Taiwan National Development Council. Business cycle indicator database. 2016. URL: http://index.ndc.gov.tw/n/en [accessed 2016-06-18] [WebCite Cache ID 6nPfnkDey]
33. Google. Google Trends. 2016. URL: https://www.google.com/trends/ [accessed 2016-06-14] [WebCite Cache ID 6d2MVXdTH]
34. Google. How Trends data is adjusted. 2016. URL: https://support.google.com/trends/answer/4365533 [accessed 2018-03-15] [WebCite Cache ID 6xw3lCki4]
35. Janz NK, Becker MH. The Health Belief Model: a decade later. Health Educ Q 1984;11(1):1-47. [doi: 10.1177/109019818401100101] [Medline: 6392204]
36. Hamilton JD, Lin G. Stock market volatility and the business cycle. J Appl Econ 1996 Sep 1;11(5):573-593. [doi: 10.1002/(SICI)1099-1255(199609)11:5<573::AID-JAE413>3.0.CO;2-T]
37. Liljeblom E, Stenius M. Macroeconomic volatility and stock market volatility: empirical evidence on Finnish data. Appl Fin Econ 1997 Aug 1;7(4):419-426. [doi: 10.1080/096031097333538]
38. Taiwan Stock Exchange Corporation. Taiwan Stock Exchange. 2016. URL: http://www.twse.com.tw/zh/page/trading/indices/MI_5MINS_HIST.html [accessed 2016-06-30] [WebCite Cache ID 5bTX5L1J9]
39. Nuti SV, Wayda B, Ranasinghe I, Wang S, Dreyer RP, Chen SI, et al. The use of google trends in health care research: a systematic review. PLoS One 2014;9(10):e109583 [FREE Full text] [doi: 10.1371/journal.pone.0109583] [Medline: 25337815]
40. Pereira MA, Kartashov AI, Ebbeling CB, Van Horn L, Slattery ML, Jacobs DR, et al. Fast-food habits, weight gain, and insulin resistance (the CARDIA study): 15-year prospective analysis. Lancet 2005;365(9453):36-42. [doi: 10.1016/S0140-6736(04)17663-0] [Medline: 15639678]
41. Todd J. United States Department of Agriculture. 2005. Changes in eating patterns and dietary quality among working-age adults, 2005-2010 URL: https://www.ers.usda.gov/webdocs/publications/45172/41995_err161.pdf?v=41775 [accessed 2018-03-28] [WebCite Cache ID 6yGcWu44n]
42. Westerterp-Plantenga MS. The significance of protein in food intake and body weight regulation. Curr Opin Clin Nutr Metab Care 2003 Nov;6(6):635-638. [doi: 10.1097/01.mco.0000098087.40916.c4] [Medline: 14557793]
43. Jeffery R, French S. Epidemic obesity in the United States: are fast foods and television viewing contributing? Am J Public Health 1998 Feb;88(2):277-280. [Medline: 9491022]
44. Harris JL, Graff SK. Protecting young people from junk food advertising: implications of psychological research for First Amendment law. Am J Public Health 2012 Feb;102(2):214-222. [doi: 10.2105/AJPH.2011.300328] [Medline: 22390435]
45. Yuan ROC. Directorate-General of Budget, Accounting and Statistics. 2016. Global reach achieving quality URL: https://www.dgbas.gov.tw/point.asp?index=2 [accessed 2018-03-15] [WebCite Cache ID 6xw3PS13P]
46. Xu X. The business cycle and health behaviors. Soc Sci Med 2013 Jan;77:126-136. [doi: 10.1016/j.socscimed.2012.11.016] [Medline: 23219162]
47. Tapia GJA. Response: on economic growth, business fluctuations, and health progress. Int J Epidemiol 2005 Dec;34(6):1226-1233 [FREE Full text] [doi: 10.1093/ije/dyi207] [Medline: 16243951]
48. Cook AR, Chen MIC, Pin LRT. Internet search limitations and pandemic influenza, Singapore. Emerg Infect Dis 2010 Oct;16(10):1647-1649 [FREE Full text] [doi: 10.3201/eid1610.100840] [Medline: 20875307]
49. Ayers J, Althouse B, Allem J, Ford D, Ribisl K, Cohen J. A novel evaluation of World No Tobacco day in Latin America. J Med Internet Res 2012 May 28;14(3):e77 [FREE Full text] [doi: 10.2196/jmir.2148] [Medline: 22634568]

Abbreviations

ARIMA: autoregressive integrated moving average
CPI: consumer price index
HBM: health belief model
RSV: relative search volumes
TWSE: Taiwan Stock Exchange

Edited by G Eysenbach; submitted 12.01.17; peer-reviewed by B Althouse, D Li, JP Allem, Q Nguyen, A Zheluk; comments to author 12.07.17; revised version received 30.11.17; accepted 12.02.18; published 06.04.18.

Please cite as:
Wang HW, Chen DR
Economic Recession and Obesity-Related Internet Search Behavior in Taiwan: Analysis of Google Trends Data
JMIR Public Health Surveill 2018;4(2):e37
URL: http://publichealth.jmir.org/2018/2/e37/
doi:10.2196/publichealth.7314
PMID:29625958

Review

Identifying Methods for Monitoring Foodborne Illness: Review of Existing Public Health Surveillance Techniques

Rachel A Oldroyd[1,2], MSc; Michelle A Morris[1,3], PhD; Mark Birkin[1,2], PhD

[1]Leeds Institute for Data Analytics, University of Leeds, Leeds, United Kingdom
[2]School of Geography, University of Leeds, Leeds, United Kingdom
[3]School of Medicine, University of Leeds, Leeds, United Kingdom

Corresponding Author:
Rachel A Oldroyd, MSc
Leeds Institute for Data Analytics
University of Leeds
Worsley Building, Level 11
Clarendon Road
Leeds, LS2 9JT
United Kingdom
Phone: 44 113 34 ext 33309
Email: r.oldroyd@leeds.ac.uk

Abstract

Background: Traditional methods of monitoring foodborne illness are associated with problems of untimeliness and underreporting. In recent years, alternative data sources such as social media data have been used to monitor the incidence of disease in the population (infodemiology and infoveillance). These data sources prove timelier than traditional general practitioner data, they can help to fill the gaps in the reporting process, and they often include additional metadata that is useful for supplementary research.

Objective: The aim of the study was to identify and formally analyze research papers using consumer-generated data, such as social media data or restaurant reviews, to quantify a disease or public health ailment. Studies of this nature are scarce within the food safety domain, therefore identification and understanding of transferrable methods in other health-related fields are of particular interest.

Methods: Structured scoping methods were used to identify and analyze primary research papers using consumer-generated data for disease or public health surveillance. The title, abstract, and keyword fields of 5 databases were searched using predetermined search terms. A total of 5239 papers matched the search criteria, of which 145 were taken to full-text review—62 papers were deemed relevant and were subjected to data characterization and thematic analysis.

Results: The majority of studies (40/62, 65%) focused on the surveillance of influenza-like illness. Only 10 studies (16%) used consumer-generated data to monitor outbreaks of foodborne illness. Twitter data (58/62, 94%) and Yelp reviews (3/62, 5%) were the most commonly used data sources. Studies reporting high correlations against baseline statistics used advanced statistical and computational approaches to calculate the incidence of disease. These include classification and regression approaches, clustering approaches, and lexicon-based approaches. Although they are computationally intensive due to the requirement of training data, studies using classification approaches reported the best performance.

Conclusions: By analyzing studies in digital epidemiology, computer science, and public health, this paper has identified and analyzed methods of disease monitoring that can be transferred to foodborne disease surveillance. These methods fall into 4 main categories: basic approach, classification and regression, clustering approaches, and lexicon-based approaches. Although studies using a basic approach to calculate disease incidence generally report good performance against baseline measures, they are sensitive to chatter generated by media reports. More computationally advanced approaches are required to filter spurious messages and protect predictive systems against false alarms. Research using consumer-generated data for monitoring influenza-like illness is expansive; however, research regarding the use of restaurant reviews and social media data in the context of food safety is limited. Considering the advantages reported in this review, methods using consumer-generated data for foodborne disease surveillance warrant further investment.

(JMIR Public Health Surveill 2018;4(2):e57) doi:10.2196/publichealth.8218

KEYWORDS

disease; review; social media; foodborne diseases; public health; infodemiology; infoveillance; digital disease detection

Introduction

Background

The Food Standards Agency (FSA) estimates that there are more than 1.7 million cases of foodborne illness contracted each year in the United Kingdom, of which 22,000 cases result in hospital admission and 700 cases result in death [1]. Defined as an infectious intestinal disease (IID), foodborne illness is caused by harmful pathogens such as parasites, viruses, and bacteria which enter the body through the ingestion of food or drink [2]. Symptoms include vomiting, diarrhea, fever, and abdominal pain and can vary from mildly uncomfortable to severe [3]. Although many victims recover in a few days, cases in vulnerable populations can incur lasting effects or result in fatality, especially if symptoms are persistent. Elderly people older than 65 years, young children under 5 years, pregnant women, and individuals with immunocompromising diseases are particularly at risk and are more likely to suffer from severe secondary effects such as dehydration. The burden of foodborne illness on the population and economy remains unacceptably high, and its reduction is a key objective in the FSA's Foodborne Disease Strategy [4].

A foodborne pathogen can infect a food vehicle at any point in the supply chain, from farm to fork; however, it can be difficult to verify foodborne illness and track an infected food vehicle unless an afflicted individual visits their general practitioner (GP) and submits a sample for laboratory testing. As GP data processing takes approximately 2 weeks, an outbreak may be escalated by delay in the identification and isolation of the responsible pathogen. GP data are not only untimely but also severely underestimate the true incidence of foodborne illness as many people choose to recover at home without visiting a medical practitioner. Combined with the infrequency of sample submissions for laboratory testing, underreporting occurs at both the patient and GP level [5]. In recent years, many studies have explored the use of online consumer generated data (CGD) to undertake public health monitoring and disease surveillance. These data, which include Twitter, restaurant reviews, and Web browser searches, are thought to have many advantages over traditional data [6]. They are timelier, may have the potential to fill gaps in the reporting process, and include additional metadata appropriate for further analysis.

Consumer-Generated Data for Disease Surveillance

Studies using CGD have ranged from influenza monitoring [5,7,8] to the surveillance of dental pain [9], but surprisingly, studies focusing on the incidence of foodborne illness are limited [10]. With the potential to improve surveillance in the food safety domain, it is important to understand and evaluate key methodologies used in CGD analysis. This review aims to identify and formally analyze primary research papers concerned with the use of CGD for disease and public health surveillance with a view to summarizing transferable methods for monitoring the outbreak and incidence of foodborne disease. It is hoped that the application of these methods may improve future policy and practice in the domain.

Methods

Review Question and Scope

Structured scoping methods were used to identify peer-reviewed papers, conference papers, and proceedings published between 2002 and 2017. Papers outlining methods concerned with, or transferable to, using CGD for the surveillance and monitoring of foodborne illness were of particular interest. CGD is defined as data created and made publically available by the general population. Public health is defined as the health of the population as a whole. Disease surveillance is defined as the monitoring of an illness or sickness presenting a set of well-defined symptoms.

Search Strategy

The abstract, title, and keyword fields of 5 individual databases were searched using predetermined search criteria. Due to the multidisciplinary nature of the review topic, the databases were specifically chosen to ensure they covered a range of discipline areas with a view to capture all relevant literature relating to disease and public health surveillance. The databases were selected to cover 3 broad topic areas: multidisciplinary (*Web of Science*); medical science (*Ovid MEDLINE*, *Embase*); and computing science (*IEEE Xplore*, *ACM*). The predetermined search terms are outlined in Table 1 and relate to 3 themes: data (eg, Twitter); application (eg, food), and methods (eg, monitor); these were adapted for each database to ensure appropriate syntax. The searches were limited to papers published after 2002 to coincide with the Web 2.0 movement. Web 2.0 describes the emergence of online communities, including the proliferation of social media, and the transition toward dynamic and user-centric Web design in the early 2000s. The search terms and the limitations detailed here were based on the methods adopted in the 2013 systematic review by Bernardo et al [10]. The full search strategy can be seen in Figure 1.

Alongside the database searches, a supplementary Google Scholar search was conducted in an attempt to capture missing literature. The search terms were *social media and infectious intestinal disease*, *restaurant review data and infectious intestinal disease*, *social media data for foodborne illness*, *novel data for foodborne illness*, *online data for food safety*, *social media and public health*, *social media and disease surveillance*, *online data and public health*, and *online data and disease surveillance*; these searches were limited to the top 100 most relevant hits. The titles of papers returned by Google Scholar and the database searches were screened for relevancy by one independent reviewer. Reference list reviews of key papers were also undertaken to ensure relevant publications were not missed. Subsequently, all relevant citations were imported into EndNote (Clarivate Analytics, Philadelphia), a reference management tool.

Table 1. Database search terms. adj4, where 2 words appear within a distance of 4 words; adj2, where the 2 words appear within a distance of 2 words. Word stems are used to ensure inflectional and derivational forms are included.

Search component	Search terms
Data	((micro-blog* or social media or twitter or yelp or trip advisor) adj4 ((public adj1 health) or influenza or (disease* adj1 surveillance)))
Application	((online or track or monitor) adj4 ((food*)or(illness*) or (gastroenteritis) or (influenza) or (infectious adj1 intestinal)))
Methods	(disease* or epidemic* or online or syndromic) adj2 (early or detect* or monitor* or model* or surveillance or control) Infoveillance
	natural adj2 (language or processing)
	Infodemiology

Figure 1. Outline of search strategy.

Textbox 1. Abstract screening exclusion criteria.

1. Studies using non-Western data, eg, Weibo or Sino microblogs
2. Studies not written in English language
3. Studies referring to disease in nonhuman populations
4. Studies concerned with the microbiological detection of disease
5. Studies concerned with public health monitoring or disease surveillance using traditional data
6. The use of social media as a tool for patient support
7. Studies conducting sentiment analysis of social media messages
8. The use of social media as a communication tool by health care organizations
9. The use of social media by researchers to disseminate medical research findings
10. Studies profiling social media users
11. Studies examining the use of mobile phone apps for infoveillance
12. Surveillance and monitoring of mental health problems and outcomes including alcoholism and suicide
13. Surveillance of drug abuse
14. Studies of smoking cessation
15. Studies concerned with noncommunicable diseases including neurological diseases, cancer, epilepsy, psychogenic seizures, migraine, and multiple sclerosis
16. Studies using search query data such as Google Flu Trends

Following deduplication, each citation deemed relevant in the previous screening stage was subject to full-text review to determine its relevance based on predetermined exclusion criteria, outlined in Textbox 1. Papers that matched the exclusion criteria were discounted at this point. Studies using CGD, including social media data and restaurant reviews for calculating the incidence of public health or disease within the population, were considered relevant. This included published journal papers, conference papers, and proceedings. Any paper not written in English language was discounted because of the absence of resources for translation.

Thematic Analysis

After full-text review, a thematic analysis was undertaken on those studies which were deemed relevant in an attempt to identify important methodological considerations. Data extraction was undertaken using a set of predefined criteria to ensure this process was standardized across each relevant study. Information relating to *topic*, *geographic region*, *primary data type*, *corpus size*, *control data type (if used)*, *keyword selection*, *methods*, *results*, *demographic analysis*, *and limitations* was extracted and summarized. See Multimedia Appendix 1 for the full data characterization form.

Results

Data Extraction

A total of 5239 papers matched the predetermined search terms during the 5 database searches. Moreover, 82 research studies were identified during the Google Scholar search and key paper reference list search. After deduplication and title and abstract screening, 145 papers were thought to discuss the use of CGD for public health and disease monitoring, and after full-text review, 62 papers were deemed relevant to this review. See Figure 2 for an overview of this process.

Data Characterization

Of 145 papers, 5 papers were systematic or scoping reviews of existent literature and 27 were qualitative overview or commentary papers discussing the strengths, challenges, and advances in novel data. In addition, 4 papers described conceptual and theoretical frameworks for the use of CGD in disease surveillance and 47 were deemed irrelevant on further inspection because of the topic, data, or methods used. A total of 62 papers proposed a process of primary CGD analysis to determine individual cases of public health or disease reporting and were therefore considered relevant. The full list of relevant papers is available in Multimedia Appendix 2.

The majority of relevant studies (40/62, 65%) described the use of CGD for monitoring outbreaks of influenza-like illness (ILI), 8 focused on the general topic of public health monitoring and looked at a spectrum of ailments such as allergies and back pain. Moreover, 7 studies discussed general disease including conjunctivitis and pertussis. Only 10 studies discussed the use of novel data in the domain of foodborne illness, gastroenteritis, or IID. Twitter data were the most common primary data source and were used in 58 of 62 studies. These studies used corpora between 1000 and 1 billion tweets. Of those studies which did not use Twitter data, 3 used Yelp restaurant reviews to explore food safety [11-13] and 1 study used internet forums [14].

The majority of studies in this review attempted to quantify disease or public health ailment incidence over a specific time interval by calculating the number of individuals reporting symptoms through via social media or through a restaurant review.

Figure 2. Search results. Many studies employed multiple methodological approaches.

Moreover, 11 of 62 studies used a basic methodological approach to calculate disease incidence, whereby the occurrence of messages containing a specific keyword or number of keywords were used to represent reports of illness. In addition, 42 studies used regression or classification techniques in an attempt to filter irrelevant messages from the data corpus, and 8 studies used unsupervised clustering-based methods to identify relevant messages. Furthermore, 15 studies used lexicon-based methods to generate statistics based on term weights and term frequencies to filter relevant messages from a large data corpus.

Thematic Analysis

A total of 4 thematic areas were identified in this review: (1) methods for calculating disease incidence using a large text corpus; (2) the challenges of working with unstructured text data; (3) the challenges of using CGD for disease surveillance; and (4) the advantages of using CGD for disease surveillance. We will discuss each theme in turn in the Discussion section of this paper.

Discussion

Methods for Calculating Disease Incidence

The methods used to calculate disease incidence are varied and wide-ranging in sophistication and complexity; therefore, with a view to discussing this theme with clarity, the methodological approaches have been divided into 4 broad classes: B) basic approach; R) regression and classification approach; C) clustering approach; and L) lexicon-based approach. This method categorization is based on a similar classification proposed by Witten and Frank [15], with the inclusion of an additional class for lexicon-based methods that do not fit into any of the previous classes. It should be noted that the process of categorizing text classification methods is not a menial task. Many methodological approaches can be extended to use different traits and are not discrete in their characteristics; they can, therefore, be considered as a member of more than one class.

The first class, basic approach, describes the least sophisticated method of disease incidence calculation. In some studies [11,16-27], simple keyword occurrence is used to calculate the incidence of disease in the population. As an example, Quincey and Kostkova [24] used a single keyword “flu” to collect messages from the Twitter application programming interface (API). Each tweet was assumed to represent a report of first-person illness, and an ILI rate was calculated based on the number of reports. Unlike other studies, this paper did not compare its results with a baseline measure of influenza, such as the rate reported through the Center for Disease Control (CDC) ILI Network; therefore, its performance cannot be assessed. However, studies using this crude approach generally report results that are highly correlated with published statistics (r>.6). In a study by Culotta [28], ILI incidence was quantified by collecting tweets that matched a small set of keywords. When compared with CDC data, the calculated rate achieved a correlation coefficient of r=.964, which suggests that lightweight approaches for disease surveillance cannot be dismissed. The main problem with the basic approach arises when the data are used for predictive purposes.

Krieck et al [29] state that online messages that include a specific disease name are more likely to be health-related communications or media papers than a report of illness. Therefore, models that calculate incidence based on a single disease keyword without adopting more sophisticated filtering techniques are extremely sensitive to false alarms. For example, the recall of a flu vaccine or a new government policy would lead a predictive model to detect a nonexistent rise in flu rates due to increased media coverage. Culotta [28] proceeded to analyze the robustness of lightweight methods against such false alarms by calculating the correlation of spurious keywords such as “vaccine” and “shot” with CDC ILINet data. The spurious keywords achieved similar correlations as nonspurious keywords, proving the need to use methods to filter false positive messages from the data corpus. Culotta [28] illustrates this by training a bag-of-words classifier to predict whether a message was reporting an ILI symptom or not. Although this did not significantly improve the model correlations, the application of this classifier reduced the mean-squared error from 0.077 to 0.023, reducing the model’s sensitivity to false alarms. This type of classifier falls within the second class of methodological approaches, R, as defined during data characterization. This class includes regression and classification techniques used to remove irrelevant messages and background noise.

Considered more sophisticated than the basic approach, these methods include probabilistic and generalized linear models and machine learning algorithms such as Support Vector Machine (SVM), Naïve Bayes (NB), and Decision Trees. These methods aim to reduce the size of the data corpus and calculate disease incidence only from messages that fit into the relevant class [30]. They are, therefore, more robust against false alarms compared with keyword-only approaches. The most commonly used method in this class, used in 22 studies [8,13,14,28,31-48], was SVM. SVM is a supervised, nonprobabilistic binary linear classifier. Provided with a labeled training dataset, SVM will learn a classification algorithm and assign unseen examples to a given class. Kang et al [13] used an SVM to label Yelp data in an attempt to predict hygiene violations for restaurants in Seattle. This study found that textual content such as unigram and bigram features, constructed of one- and two-word terms, respectively, are able to predict health violations with high accuracy (83%). Textual features outperformed measures such as review rating and inspection history, which achieved accuracies of 53% and 72%, respectively. Kang et al [13] found that terms such as *student*, *door*, and *the size* frequently occurred in restaurants with low hygiene scores, whereas terms referring to *selection*, *atmosphere*, and *ambiance* were indicative of a hygienic restaurant. This study suggests that factors contributing to food safety concerns can be extracted from restaurant review site messages, and highlights the capacity of text reviews as a useful indicator of food hygiene practices.

In a similar study, Kate et al [14] used SVM alongside a multinomial NB classifier to monitor food safety violation reports from internet forums. NB is the second most commonly used classifier in this methodological class, used in 13 studies [5,8,14,31,38,43,47-53]. It is a supervised classification algorithm that is probabilistic in its approach. NB, therefore, assigns new examples to a given class based on a calculated degree of certainty. When applied to the problem of filtering relevant and irrelevant messages relating to food safety violations, Kate et al [14] found that the NB model was outperformed by the SVM classifier, which achieved precision and recall values of 0.795 and 0.75, respectively. Precision and recall values for the NB model were not presented in this study; however, in comparative studies undertaken by Achrekar [5] and Carlos and Matos [31], SVM was found to achieve higher precision and recall values than NB. Although these studies suggest that SVM may be a superior classification compared with NB, it should be noted that both models are sensitive to parameter optimization. Different parameters can drastically change the results of a model; consequently, NB may perform better in a different scenario, given a different dataset and different parameters. A possible explanation for NB’s inferior performance in these particular studies is the way in which it considers terms located in the same message. When making classification decisions, SVM takes into consideration the correlations between single-term values that constitute a message, whereas NB assumes that each term contributes independently to the probability that the message is relevant or not, without considering interterm correlations [54]. Subsequently, it may be that terms within messages relating to public health and disease are more interrelated than terms in messages relating to other domains. In this case, SVM would be a more suitable technique than NB. Despite NB’s underperformance in these particular studies, one of its advantages is that it requires only a small number of training data examples to estimate the required parameters. The requirement of quality training data is considered the main limitation of all supervised text-based classification methods.

Studies using methods that fall within the second class, regression and classification, report the highest correlations with baseline measures; however, before classification can begin, they require training. Achrekar [5] used a rule-based classifier and achieved a correlation of r=.98, and similarly, Doan [55] used a hybrid classifier using both semantic and textual inputs and achieved a correlation of r=.98, but the performance of these supervised algorithms is greatly dependent on the training data. Creation of a training dataset is not a simple task and can be extremely resource-intensive [56]. Ideally, a training dataset should be representative of the real-world problem, of sufficient size to capture input-to-input and input-to-output feature relationships, and should be composed of independent examples [57]. Most studies identified in this review generated training data by manually labeling a small sample of messages [32,49], but some studies [5,33-35,50] leveraged virtual human intelligence via Amazon Mechanical Turk (AMT) for the task. AMT provides a scalable and on-demand workforce, allowing large training datasets to be generated in a less resource-intensive way. Using AMT, Achrekar [5] labeled 25,000 tweets and trained an SVM to classify relevant and irrelevant tweets relating to ILI. The provision of such a large training dataset could explain the high correlations (r=.98) reported by this study. Not only was AMT used to create large training datasets but it was also explored by some studies as a utility to overcome class imbalance.

Document classifiers work best when the number of messages deemed relevant and irrelevant is approximately equal. When this is not the case, eg, when only 5% of messages report foodborne illness, the classifier is biased toward the majority class in an attempt to minimize error scores. This problem is known as class imbalance. In an attempt to address class imbalance, Sadilek et al [35] used a method of human-guided machine learning, whereby instances belonging to the minority class were actively provided to the model during the training process. This study used AMT to find reports of foodborne illness, rewarding the workforce for each unique instance they found. This model achieved a precision value of 0.63 compared with data provided by the Department of Health and Mental Hygiene; however, the training dataset constituted only 200 examples. It is thought a higher precision value may have been achieved from a larger training dataset that was more representative of the testing data. Schomberg et al [58] used an alternative method to combat class imbalance. This study found that Chinese restaurants were more likely to have health code violations compared with other restaurants (25% vs 7%). Yelp reviews from Chinese restaurants were, therefore, used to train a predictive model as they were thought to contain more instances of the positive, minority class. The authors hoped to reduce the number of false positives by training the model in favorable conditions and were able to detect health violations in 78% of restaurants in the pilot study. Alongside the problems of class imbalance, many studies also discussed the challenges and importance of using discrete datasets for the training and testing process [31,37]. If overlapping data are used for both the training and testing process, the model will label data with which it was trained, overestimating its performance. Therefore, a data corpus should be large enough to split data into distinct training and testing phases.

The third methodological approach outlined in this review was class C clustering. This class outlines models that aim to identify hidden groupings and patterns within a data corpus. Clustering algorithms maximize the similarity of messages within a specific class while ensuring messages are as distinct as possible from those assigned to other classes. Many clustering models are semisupervised or unsupervised and are therefore less resource-intensive than supervised classification models, and their performance is not dependent on the provision of quality training data. Methodological approaches in this class include k-Nearest Neighbor (k-NN), Markov-Chain State modeling, and Latent Dirichlet Allocation. A total of 8 studies in this review adopted clustering techniques to filter hidden states from the text corpus [8,14,20,31,38,59-61]. Of these methods, the k-NN algorithm is considered one of the simplest machine learning algorithms as the function is approximated locally, based on the terms closest neighbors, and all computation is deferred until classification. Nargund and Natarajan [38] used a minimally supervised k-NN alongside SVM and NB to identify messages reporting first-person allergies and messages discussing allergies. The algorithm was able to determine different types of allergy, including milk, peanut, and dog allergy, and outperformed NB and SVM with reported precision and recall values of 0.864 and 0.852, respectively. Conversely, in a study to determine ILI incidence using Portuguese tweets, Carlos and Matos [31] report that SVM achieved higher precision and recall values than k-NN. An explanation of this discrepancy may lie within the nature of the classification problems. Classifying ILI tweets can be considered dichotomous as the output will belong to either the relevant or irrelevant class. As this is a binary classification task, it makes sense that SVM performs well. Alternatively, the k-NN algorithm performs best when identifying and assigning tweets to multiple hidden classes and is, therefore, more appropriate for the allergy problem outlined in the study by Nargund and Natarajan [38].

Finally, the fourth methodological approach identified in this review relates to lexicon-based approaches, class L. This class describes methods including word embeddings, term statistics, and frequent pattern mining, whereby statistics are generated based on the frequency or relative importance of a term in relation to a topic. By considering the terms that constitute a message, these models rank messages based on their overall significance. A total of 15 studies used lexicon-based methods to calculate disease incidence [14,33,34,36-38,43,44, 52,58,62-66]. For example, Velardi [62] proposed a model for the early detection of epidemics. This study weighted clusters of naïve and medical terms and assigned them to distinct classes based on their topics, for example, *cold* or *allergy*. When used to calculate rates of ILI, this model was able to achieve an extremely high correlation coefficient of r=.998 compared with CDC ILINet data. In another lexicon-based study, Zou et al [63] used a deep learning approach to investigate rates of IID via social media. Topical vocabulary was identified by calculating a similarity score between all word embeddings in the data corpus and the word embeddings of terms that describe IID symptoms such as *vomit* and *diarrhea*. A word embedding is defined as the words surrounding a context term. Word embeddings were ranked according to their similarity score, and those with the highest score were used to define the IID vocabulary. Zou et al [63] applied the keywords in a linear Elastic Net regression and a nonlinear Gaussian Process covariance function in an attempt to model nonlinearities between the keyword frequencies and the baseline measure of IID, as reported by Public Health England (PHE) over set time intervals. The Gaussian Process covariance function outperformed the Elastic Net regression reporting correlations of up to r=.77. Compared with the study undertaken by Velardi [62], this model reports relatively low correlation coefficients. However, this may be because the gold standard measure against which the IID correlations are calculated is not as representative as the gold standard measure for ILI in Velardi [62]. IID is notoriously hard to diagnose, and PHE data are based solely on laboratory-confirmed cases of pathogens. As IID is severely underreported at the patient and GP level, PHE data are not representative of the true incidence of disease. In comparison, the CDC collects data from more than 400 public health and clinical laboratories to calculate ILI rates and therefore may be considered more accurate as a baseline measure. Lexicon-based approaches are highly dependent on both the size of the vocabularies used and the similarity score threshold value beyond which the word embeddings are no longer deemed relevant to the defined topic; therefore, an alternative vocabulary and threshold value may yield differing results.

Working With Unstructured Text Data

Although some studies used datasets from previous studies, eg, Doan et al [55] used the dataset collected by Culotta [67], many collected primary data and therefore faced the methodological challenge of preprocessing online messages into a useable format. Many studies used natural language processing (NLP) methods to remove HTML characters, emoticons, stop words, and punctuation in an attempt to filter noise from the useful part of the message. Stemming and lemmatization techniques were also used to reduce inflectional and derivational forms of a word to a common word base. Popular libraries for NLP included the Natural Language Toolkit [31,51], TextBlob [68] in Python, and the Apache OpenNLP library [64]. Alongside removing useless noise from the message, these preprocessing techniques also ensure that the data corpus is in a useful form for subsequent phases of analysis, such as the application of a document classifier.

In an attempt to filter spurious messages such as health communications and media-related tweets before disease incidence calculation, many studies removed retweets, replies, and tweets with a URL. As mentioned previously, these messages are unlikely to represent first-person accounts of disease and can increase the model's sensitivity to false alarms. To illustrate this, Aslam et al [32] observed the correlation coefficients of 4 Twitter datasets against ILI rates published by the United States CDC—nonretweets, retweets, tweets with a URL, and tweets without a URL. Although the results differed spatially, a general trend observed was that nonretweets and tweets without a URL provided higher and more statistically significant correlations (r>.5) in comparison with the other 2 categories, reaffirming that retweets and tweets with a URL are not likely to represent a report of illness. Alongside removing retweets and tweets with a URL, many studies also discussed the importance of using feature selection either on data collection or as an initial filtering technique before more complex methods were undertaken. Feature selection reduces the size of the data corpus and attempts to remove messages that are highly likely to be irrelevant.

For feature selection, many studies selected only tweets that matched a keyword list of relevant terms, built in various ways. Some consulted experts in the field to generate a list of terms relating to disease symptoms [39,60], whereas others mined blogs and websites to collect terms [55,69,70]. To identify symptom-related words, Velardi et al [62] collected naïve and technical keyword pairs describing infectious disease from Google and Wikipedia. Similarly, Doan et al [55] collected syndrome terms from the BioCaster Ontology database [71], a knowledge model of layman terms. Other studies used document seeding to generate relevant keyword lists with which to select features. This process involves collecting a seed set of online messages matching 1 or 2 keywords and further expanding the keyword list to include the seed set's most frequently occurring words. Chen et al [59] used 230 keywords defined by Chakraborty [72] and in-house experts to retrieve the seed set; the keyword set was then expanded to 2739 frequently occurring words for subsequent phases of data collection. Similarly, Culotta [7] retrieved a candidate set of tweets matching 4 keywords: *cough*, *sore throat*, *headache*, and *flu* and then selected the top 5000 frequently occurring terms. This study compared the performance of both residual sum of squares (RSS) and Pearson correlation coefficient for candidate seeding, and found that RSS performed the best. Other techniques include using the chi-square test to identify relevant trends for feature selection, adopted by Chew and Eysenbach [17], and selection based on geographical location [32,35,40,73].

Challenges of Using Consumer-Generated Data for Disease Surveillance

The reduction of false positives and removal of spurious messages was the main methodological challenge reported by the majority of studies in this review. Although it was generally reported that high correlations against calculated results and published statistics could be achieved with a fairly crude model, these models are sensitive to increased media coverage and, therefore, prone to false alarms if used for predictive purposes [28]. Models that reported high performance and robustness against false alarms used more sophisticated methods of document classification, particularly those methods assigned to the regression and classification class (R). Although the performance of classification and regression models is highly dependent on the provision of quality training data, the collection of which can be resource intensive, they are able to achieve extremely high correlations against baseline measures compared with a basic approach, a clustering approach or a lexicon-based approach.

Related to the challenges associated with reducing false positives is the process of dealing with sarcastic and ironic messages.

Greaves et al [74] state that “Sarcasm and irony, a feature of the British and US cultures, are almost impossible to process”; however, this does not negate the potential for studies using CGD for public health and disease surveillance. Indeed, a model that assumes term independence and does not consider interterm relationships such as NB is more likely to wrongly classify a sarcastic or ironic message. Alternatively, models that use a holistic approach and consider interterm relationships are better equipped to deal with sarcasm, and many methodological approaches have been proposed to deal with this problem. These methods include pattern-based approaches [75], hashtag tokenizers [76], and context incongruity [77].

A further methodological limitation of using CGD for disease surveillance is demographic representativeness. As certain demographic groups, such as elderly people, are less likely to use the internet, they are underrepresented in data derived from social media and review sites. Although this limitation is well discussed in the literature, only 8 of 62 relevant studies mentioned or undertook demographic analysis. Aslam et al [32] state that 31% of Twitter users are aged between 18 and 29 years. Broniatowski et al [41] and Carlos and Matos [31] also discussed how this age group are well represented compared with other users. Achrekar et al [5] carried out age-based ILI prediction on a small sample and found that the proposed model achieved high correlations with outbreaks among 5- to 49-year-olds but did not represent other age groups as well. These findings reaffirm that younger age groups are more prevalent on Twitter. Alternatively, Culotta [7] states the following:

> *...despite the fact that Twitter appears targeted to a young demographic, it in fact has quite a diverse set of users. The majority of Twitter’s nearly 10 million unique visitors in February 2009 were 35 years or older, and a nearly equal percentage of users are between ages 55 and 64 as are between 18 and 24.*

There is no clear agreement on the subject, and further work is required to explore the demographic representativeness of social media and review datasets and understand the effect this has on the accuracies of models such as those discussed in this review.

Advantages of Consumer-Generated Data

Using CGD to calculate disease incidence and public health ailments has certain advantages over traditional datasets. CGD often contains additional metadata and text, which is not available in traditional data. When writing a restaurant review, a consumer may comment on the cleanliness of the restaurant, the service, and the food they ate, providing valuable information relating to food safety procedures and the restaurant environment which can be used to inform food safety research [13]. Schomberg et al [58] used ~70,000 Yelp reviews from San Francisco to predict restaurants’ likelihood of health violation. This study labeled each review as 1 of the 3 categories depending on keyword matches: physical environment match; sentiment match; and foodborne illness match. Using the additional information in this way provided insight into other aspects of the restaurant experience, which may be helpful to health inspectors. In a similar study, Nsoesie et al [11] extracted specific food vehicles from online restaurant reviews reporting foodborne illness and ranked them in order of frequency. This study found a high correlation (.78) between the frequency of food vehicles reported in restaurant reviews and frequently occurring food vehicles in the CDC’s Foodborne Outbreak Online Database. Although not explicitly concerned with identifying cases of foodborne illness, this study outlines the importance of additional information provided via online reviews in tracking harmful pathogens in the supply chain.

Another advantage reported in almost each study was the timeliness of novel data compared with traditional data. Traditionally, public health monitoring is undertaken using GP data reported via national surveillance, which has a latency of around 2 weeks between GP appointment and data publication [78]. Due to this latency, the data are of limited use for monitoring public health outbreaks [5]. Although GP data take around 2 weeks to publish, data collected through social media or online reviews can be collected in near real time, providing a valuable resource for the timely identification and isolation of a foodborne pathogen. In addition, as many people choose to remain at home and recover from foodborne illness without visiting their GP, reports of foodborne illness identified through social media and online reviews may help to fill gaps in national surveillance data caused by underreporting at both the GP and patient level.

Conclusions

This review identified and formally analyzed 62 primary research papers concerned with the use of CGD for public health monitoring and disease surveillance. The methodological approaches adopted by these studies were categorized into 4 broad categories: B) basic approach; R) regression and classification approaches; C) clustering approaches; and L) lexicon-based approaches and were analyzed with a view to understanding their strengths, weaknesses, and application in the domain of food safety. Only 10 research studies that used methods for monitoring foodborne illness or IID were identified. However, the methods adopted by other studies are highly transferable to the surveillance of foodborne illness, and many recommendations have emerged through the analysis of these methods.

Studies that achieved the highest and most significant correlations against published statistics adopted supervised machine learning document classifiers, the most common of which was SVM. Although the performance of document classifiers depends highly on the application and input parameters, SVM was found to be highly suitable for binary classification tasks, whereby the output is dichotomous. This includes tasks such as classifying positive and negative reports of foodborne illness. Studies using a classifier to filter false positives were found to be more robust against false alarms than studies adopting a basic approach based on keyword incidence. Feature selection was also found to improve the performance of the model by removing messages deemed unlikely to be relevant before classification. Of the feature selection techniques, filtering messages using symptom-specific keyword lists based on existing knowledge mined from blogs and websites was the most suitable. This type of keyword list was more likely to retrieve messages reporting illness compared with disease-specific keywords such as “food poisoning.”

The demographic limitations of CGD are unclear, and future work should focus on understanding the effect of these limitations on model outcomes. Demographic limitations were only discussed in a handful of reviews. However, provisional findings show that people aged between 18 and 29 years are well represented on Twitter but are underrepresented in national foodborne illness outbreak statistics, as they prefer to recover at home without seeking medical advice from their GP. This highlights the utility of CGD to complement traditional data sources. The lack of primary research in the area of CGD for food safety provides a strong case for further research. Considering the reported success of studies in other health-related fields, it is thought that CGD could prove useful in helping to inform and improve current inspection procedures in the United Kingdom by identifying problematic restaurants and specific outbreaks of disease. In the long term, a model that can successfully detect reports of foodborne illness through social media data and online restaurant reviews could reduce the burden on the economy and, more importantly, the population. CGD may also have the capacity to fill gaps in national surveillance data and combat problems associated the underestimation of disease incidence.

Acknowledgments

The authors wish to acknowledge funding from the Food Standards Agency and the Economic and Social Research Council (grant number ES/J500215/1).

Authors' Contributions

RAO carried out database and Google Scholar searches, screened titles and abstracts against the inclusion and exclusion criteria, characterized the results, and authored the paper. MAM and MB edited the paper and provided guidance on the creation of the review protocol.

Conflicts of Interest

MAM is an inventor and shareholder at Dietary Assessment Ltd, a University of Leeds spin-out company.

Multimedia Appendix 1

Data characterization form used to extract relevant information during full-text review.

[PDF File (Adobe PDF File), 10KB - publichealth_v4i2e57_app1.pdf]

Multimedia Appendix 2

Results of data characterization and methods coding (Alco: alcohol sales; D: disease; FBI: foodborne Illness; ILI: influenza-like illness; IID: infectious intestinal disease; PH: public health). Coding: ARX: autoregressive modeling with exogenous terms; BOLASSO: bootstrapped least absolute shrinkage and selection operator; BOW: bag of words; DT: decision tree; k-NN: K-nearest neighbor; LASSO: least absolute shrinkage and selection operator: LDA: latent dirichlet allocation; NB: naïve Bayes; PDE: partial differential equation; POS: part of speech tagging; RF: random forest; SVM: support vector machine; TS: term statistics).

[PDF File (Adobe PDF File), 162KB - publichealth_v4i2e57_app2.pdf]

References

1. Tam C, Larose T, O'Brien S, Adak B, Cowden J, Evans M, et al. Food.gov.uk. 2014. Costed extension to the Second Study of Infectious Intestinal Disease in the Community: Identifying the proportion of foodborne disease in the UK and attributing foodborne disease by food commodity URL: https://www.food.gov.uk/sites/default/files/IID2%20extension%20report%20-%20FINAL%2025%20March%202014_0.pdf [accessed 2017-06-15] [WebCite Cache ID 6xuaXor8j]
2. World Health Organisation. 2015. Food Safety Fact Sheet URL: http://www.who.int/mediacentre/factsheets/fs399/en/ [accessed 2017-06-16] [WebCite Cache ID 6rEnV8Wc8]
3. American Medical Association, American Nurses Association--American Nurses Foundation, Centers for Disease Control and Prevention, Center for Food Safety and Applied Nutrition, Food and Drug Administration, Food Safety and Inspection Service, US Department of Agriculture. Centers for Disease Control and Prevention. 2004. Diagnosis and management of foodborne illnesses URL: https://www.cdc.gov/mmwr/preview/mmwrhtml/rr5304a1.htm [WebCite Cache ID 6wvny1RJq]
4. Food Standards Agency. Food.gov.uk. Foodborne Disease Strategy 2010-15: An FSA programme for the reduction of foodborne disease in the UK URL: https://www.food.gov.uk/sites/default/files/multimedia/pdfs/fds2015.pdf [accessed 2017-06-11] [WebCite Cache ID 6rEnppzBT]
5. Achrekar H, Gandhe A, Lazarus R, Yu S, Liu B. Twitter Improves Seasonal Influenza Predictions. 2012 Presented at: International Conference in Health Informatics; February 1-4, 2012; Algarve, Portugal.

6. Zhao L, Chen J, Chen F, Wang W, Lu CT, Ramakrishnan N. SimNest: social media nested epidemic simulation via online semi-supervised deep learning. Proc IEEE Int Conf Data Min 2015 Nov;2015:639-648 [FREE Full text] [doi: 10.1109/ICDM.2015.39] [Medline: 27453696]
7. Culotta A. Towards detecting influenza epidemics by analyzing Twitter messages. 2010 Presented at: The first workshop on social media analytics; July 25-28, 2010; Washington, DC, USA p. 115-122. [doi: 10.1145/1964858.1964874]
8. Aramaki E, Maskawa S, Morita M. Twitter catches the flu: detecting influenza epidemics using Twitter. 2011 Presented at: 2011 conference on empirical methods in natural language processing; July 27-31, 2011; Edinburgh, UK p. 1568-1576.
9. Heaivilin N, Gerbert B, Page JE, Gibbs JL. Public health surveillance of dental pain via Twitter. J Dent Res 2011 Sep;90(9):1047-1051 [FREE Full text] [doi: 10.1177/0022034511415273] [Medline: 21768306]
10. Bernardo TM, Rajic A, Young I, Robiadek K, Pham MT, Funk JA. Scoping review on search queries and social media for disease surveillance: a chronology of innovation. J Med Internet Res 2013;15(7):e147 [FREE Full text] [doi: 10.2196/jmir.2740] [Medline: 23896182]
11. Nsoesie EO, Kluberg SA, Brownstein JS. Online reports of foodborne illness capture foods implicated in official foodborne outbreak reports. Prev Med 2014 Oct;67:264-269 [FREE Full text] [doi: 10.1016/j.ypmed.2014.08.003] [Medline: 25124281]
12. Harrison C, Jorder M, Stern H, Stavinsky F, Reddy V, Hanson H, et al. CDC.: CDC (Morbidity and mortality weekly report); 2014. Using online reviews by restaurant patrons to identify unreported cases of foodborne illness - New York City, 2012-2013 URL: https://www.cdc.gov/MMWr/preview/mmwrhtml/mm6320a1.htm [accessed 2017-06-13] [WebCite Cache ID 6wvqBmzTU]
13. Kang JS, Kuznetsova P, Choi Y, Luca M. Harvard Business School. Cambridge, Massachusetts: Harvard Business School; 2013. Using Text Analysis to Target Government Inspections: Evidence from Restaurant Hygiene Inspections and Online Reviews URL: http://www.hbs.edu/faculty/Publication%20Files/Luca_Inspections_328fbdfc-4cbe-4b8e-8549-3a21c7a56279.pdf [accessed 2017-06-15] [WebCite Cache ID 6rEwdT6kB]
14. Kate K, Negi S, Kalagnanam J. Monitoring food safety violation reports from internet forums. Stud Health Technol Inform 2014;205:1090-1094. [Medline: 25160357]
15. Witten IH, Frank E, Hall MA. Data Mining: Practical Machine Learning Tools and Techniques with Java Implementations (The Morgan Kaufmann Series in Data Management Systems). San Francisco: Morgan Kaufman; 2005.
16. Achrekar H, Gandhe A, Lazarus R, Yu S, Liu B. Predicting flu trends using Twitter data. 2011 Presented at: The First International Workshop on Cyber Networking Systems; July 10-15, 2011; Shanghai. [doi: 10.1109/INFCOMW.2011.5928903]
17. Chew C, Eysenbach G. Pandemics in the age of Twitter: content analysis of tweets during the 2009 H1N1 outbreak. PLoS One 2010;5(11):e14118 [FREE Full text] [doi: 10.1371/journal.pone.0014118] [Medline: 21124761]
18. Chorianopoulos K, Talvis K. Flutrack.org: open-source and linked data for epidemiology. Health Inform J 2016;22(4):962-974. [doi: 10.1177/1460458215599822] [Medline: 26351261]
19. Deiner MS, Lietman TM, McLeod SD, Chodosh J, Porco TC. Surveillance tools emerging from search engines and social media data for determining eye disease patterns. J Am Med Assoc Ophthalmol 2016 Sep 01;134(9):1024-1030 [FREE Full text] [doi: 10.1001/jamaophthalmol.2016.2267] [Medline: 27416554]
20. Diaz-Aviles E, Stewart A. Tracking Twitter for Epidemic Intelligence. In: Proceedings of the 4th Annual ACM Web Science Conference. 2012 Presented at: Annual ACM Web Science Conference; June 22-24, 2012; Evanston, Illinois p. 82-85.
21. Gesualdo F, Stilo G, Agricola E, Gonfiantini MV, Pandolfi E, Velardi P, et al. Influenza-like illness surveillance on Twitter through automated learning of naïve language. PLoS One 2013;8(12):e82489 [FREE Full text] [doi: 10.1371/journal.pone.0082489] [Medline: 24324799]
22. Harris JK, Mansour R, Choucair B, Olson J, Nissen C, Bhatt J, et al. Health department use of social media to identify foodborne illness - Chicago, Illinois, 2013-2014. MMWR Morb Mortal Wkly Rep 2014 Aug 15;63(32):681-685 [FREE Full text] [Medline: 25121710]
23. Nagel AC, Tsou MH, Spitzberg BH, An L, Gawron JM, Gupta DK, et al. The complex relationship of realspace events and messages in cyberspace: case study of influenza and pertussis using tweets. J Med Internet Res 2013 Oct 24;15(10):e237 [FREE Full text] [doi: 10.2196/jmir.2705] [Medline: 24158773]
24. de Quincey E, Kostkova P. Early warning and outbreak detection using social networking websites: the potential of Twitter. In: Electronic Healthcare. eHealth 2009. Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering, vol 27. Berlin, Heidelberg: Springer; 2010:978-973.
25. Yom-Tov E, Borsa D, Cox IJ, McKendry RA. Detecting disease outbreaks in mass gatherings using Internet data. J Med Internet Res 2014 Jun 18;16(6):e154 [FREE Full text] [doi: 10.2196/jmir.3156] [Medline: 24943128]
26. Barros J. Text Mining from Social Media for Public Health Applications. 2017 Presented at: International Conference on Digital Health; July 2-5, 2017; London, United Kingdom. [doi: 10.1145/3079452.3079475]
27. Wang F, Wang H, Xu K, Raymond R, Chon J, Fuller S, et al. Regional level influenza study with geo-tagged Twitter data. J Med Syst 2016 Aug;40(8):189. [doi: 10.1007/s10916-016-0545-y] [Medline: 27372953]
28. Culotta A. Lightweight methods to estimate influenza rates and alcohol sales volume from Twitter messages. Lang Resour Eval 2012;47(1):217-238. [doi: 10.1007/s10579-012-9185-0]
29. Krieck M, Dreesman J, Otrusina L, Denecke K. A New Age of Public Health: Identifying Disease Outbreaks by Analyzing Tweets. 2011 Presented at: 3rd international ACM conference on web science; June 14-17, 2011; Koblenz, Germany.

30. Hirose H, Wang L. Prediction of Infectious Disease Spread Using Twitter: A case of influenza. 2012 Presented at: 2012 Fifth International Symposium on Parallel Architectures, Algorithms and Programming; Dec 17-20, 2012; Taipei, Taiwan.
31. Carlos J, Matos S. Predicting Flu Incidence from Portuguese Tweets. 2013 Presented at: International Work-Conference on Bioinformatics and Biomedical Engineering; March 18-20, 2013; Grenada, Spain.
32. Aslam AA, Tsou M, Spitzberg BH, An L, Gawron JM, Gupta DK, et al. The reliability of tweets as a supplementary method of seasonal influenza surveillance. J Med Internet Res 2014;16(11):e250 [FREE Full text] [doi: 10.2196/jmir.3532] [Medline: 25406040]
33. Paul MJ, Dredze M. Discovering health topics in social media using topic models. PLoS One 2014 Aug;9(8):e103408 [FREE Full text] [doi: 10.1371/journal.pone.0103408] [Medline: 25084530]
34. Paul M, Dredze M. A model for mining public health topics for twitter. 2011 Presented at: 5th International AAAI Conference on Web blogs and Social Media; July 17-21, 2011; Barcelona, Spain.
35. Sadilek A, Brennan S, Kautz H, Silenzio V. Which Restaurants Should You Avoid Today? 2013 Presented at: First AAAI Conference on Human Computation and Crowdsourcing; November 7-9, 2013; Palm Springs, California.
36. Paul M, Dredze M. You Are What You Tweet: Analyzing Twitter for Public Health. 2011 Presented at: 5th International AAAI Conference on Web blogs and Social Media; July 17-21, 2011; Barcelona, Spain.
37. Allen C, Tsou MH, Aslam A, Nagel A, Gawron J. Applying GIS and machine learning methods to Twitter data for multiscale surveillance of influenza. PLoS One 2016;11(7):e0157734 [FREE Full text] [doi: 10.1371/journal.pone.0157734] [Medline: 27455108]
38. Nargund K, Natarajan S. Public Health Allergy Surveillance Using Micro-blogs. 2016 Presented at: International Conference on Advances in Computing, Communications and Informatics; Sept 21-24, 2016; Jaipur, India. [doi: 10.1109/ICACCI.2016.7732248]
39. Denecke K, Krieck M, Otrusina L, Smrz P, Dolog P, Nejdl W, et al. How to exploit Twitter for public health monitoring? Methods Inf Med 2013;52(4):326-339. [doi: 10.3414/ME12-02-0010] [Medline: 23877537]
40. Sadilek A, Kautz H, DiPrete L, Labus B, Portman E, Teitel J, et al. Deploying nEmesis: Preventing Foodborne Illness by Data Mining Social Media. 2016 Presented at: Fourth AAAI Conference on Human Computation and Crowdsourcing; Oct 30-Nov 3, 2016; Phoenix, Arizona, USA.
41. Broniatowski DA, Paul MJ, Dredze M. National and local influenza surveillance through Twitter: an analysis of the 2012-2013 influenza epidemic. PLoS One 2013 Dec;8(12):e83672 [FREE Full text] [doi: 10.1371/journal.pone.0083672] [Medline: 24349542]
42. Hartley DM, Giannini CM, Wilson S, Frieder O, Margolis PA, Kotagal UR, et al. Coughing, sneezing, and aching online: Twitter and the volume of influenza-like illness in a pediatric hospital. PLoS One 2017 Jul;12(7):e0182008 [FREE Full text] [doi: 10.1371/journal.pone.0182008] [Medline: 28753678]
43. Lee K, Agrawal A, Choudhary A. Mining Social Media Streams to Improve Public Health Allergy Surveillance. 2015 Presented at: Proceedings of the IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining; August 25-28, 2015; Paris, France. [doi: 10.1145/2808797.2808896]
44. Parker J, Wei Y, Yates A, Frieder O, Goharian N. A framework for detecting public health trends with Twitter. 2013 Presented at: IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining; August 25-28, 2013; Niagra, Ontario, Canada. [doi: 10.1145/2492517.2492544]
45. Santillana M, Nguyen AT, Dredze M, Paul MJ, Nsoesie EO, Brownstein JS. Combining search, social media, and traditional data sources to improve influenza surveillance. PLoS Comput Biol 2015 Oct;11(10):e1004513 [FREE Full text] [doi: 10.1371/journal.pcbi.1004513] [Medline: 26513245]
46. Sofean M, Smith M. A real-time disease surveillance architecture using social networks. Stud Health Technol Inform 2012;180:823-827. [Medline: 22874307]
47. Thapen N, Simmie D, Hankin C, Gillard J. DEFENDER: detecting and forecasting epidemics using novel data-analytics for enhanced response. PLoS One 2016 May;11(5):e0155417 [FREE Full text] [doi: 10.1371/journal.pone.0155417] [Medline: 27192059]
48. Zuccon G, Khanna S, Nguyen A, Boyle J, Hamlet M, Cameron M. Automatic detection of tweets reporting cases of influenza like illnesses in Australia. Health Inf Sci Syst 2015;3(Suppl 1 HISA Big Data in Biomedicine and Healthcare 2013 Con):S4 [FREE Full text] [doi: 10.1186/2047-2501-3-S1-S4] [Medline: 25870759]
49. Khan MA, Iwai M, Sezaki K. A robust and scalable framework for detecting self-reported illness from twitter. 2012 Presented at: IEEE 14th International Conference on e-Health Networking, Applications and Services; October 10-13, 2012; Beijing, China. [doi: 10.1109/HealthCom.2012.6379425]
50. Yin Z, Fabbri D, Rosenbloom ST, Malin B. A scalable framework to detect personal health mentions on Twitter. J Med Internet Res 2015;17(6):e138 [FREE Full text] [doi: 10.2196/jmir.4305] [Medline: 26048075]
51. Dai X, Bikdash M. Hybrid Classification for Tweets Related to Infection with Influenza. 2015 Presented at: IEEE Southeastcon; April 9-12, 2015; Fort Lauderdale, Florida p. 1-5. [doi: 10.1109/SECON.2015.7133015]
52. Dai X, Bikdash M, Meyer B. From social media to public health surveillance: word embedding based clustering method for twitter classification. 2017 Presented at: IEEE SouthEast Con; March 30-April 02; Charlotte, NC, USA. [doi: 10.1109/SECON.2017.7925400]

53. Santos JC, Matos S. Analysing Twitter and web queries for flu trend prediction. Theor Biol Med Model 2014 May 07;11 Suppl 1:S6 [FREE Full text] [doi: 10.1186/1742-4682-11-S1-S6] [Medline: 25077431]
54. Sagale A, Kale S. Combining naive Bayesian and support vector machine for intrusion detection system. Int J Comput Technol 2014;1(3):61-65.
55. Doan S, Ohno-Machado L, Collier N. Enhancing Twitter Data Analysis with Simple Semantic Filtering: Example in Tracking Influenza-Like Illnesses. 2012 Presented at: IEEE Second International Conference on Healthcare Informatics, Imaging and Systems Biology; September 27-28, 2012; California, USA p. 62-71 URL: https://doi.org/10.1109/HISB.2012.21 [doi: 10.1109/HISB.2012.21]
56. Lim S, Tucker CS, Kumara S. An unsupervised machine learning model for discovering latent infectious diseases using social media data. J Biomed Inform 2017 Dec;66:82-94 [FREE Full text] [doi: 10.1016/j.jbi.2016.12.007] [Medline: 28034788]
57. Ratner A, De Sa C, Wu S, Selsam D, Ré C. Data Programming: Creating Large Training Sets, Quickly. 2016 Presented at: 30th Conference on Neural Information Processing Systems; December 5-10, 2016; Barcelona, Spain.
58. Schomberg JP, Haimson OL, Hayes GR, Anton-Culver H. Supplementing public health inspection via social media. PLoS One 2016 Mar;11(3):e0152117 [FREE Full text] [doi: 10.1371/journal.pone.0152117] [Medline: 27023681]
59. Chen LZ, Tozammel Hossain KS, Butler P, Ramakrishnan N, Prakash BA. Syndromic surveillance of Flu on Twitter using weakly supervised temporal topic models. Data Min Knowl Disc 2015 Sep 8;30(3):681-710. [doi: 10.1007/s10618-015-0434-x]
60. Chen L, Tozammel Hossain KS, Butler P, Ramakrishnan N, Prakash B. Flu Gone Viral: Syndromic Surveillance of Flu on Twitter Using Temporal Topic Models. 2014 Presented at: IEEE International Conference on Data Mining; December 14-17, 2014; Shenzen, China. [doi: 10.1109/ICDM.2014.137]
61. Chon J, Raymond R, Wang HY, Wang F. Modeling Flu Trends with Real-Time Geo-tagged Twitter Data Streams. 2015 Presented at: 10th International Conference on Wireless Algorithms, Systems, and Applications; August 20-22, 2015; Qufu, China p. 60-69.
62. Velardi P, Stilo G, Tozzi AE, Gesualdo F. Twitter mining for fine-grained syndromic surveillance. Artif Intell Med 2014 Jul;61(3):153-163. [doi: 10.1016/j.artmed.2014.01.002] [Medline: 24613716]
63. Zou B, Lampos V, Gorton R, Cox JI. On Infectious Intestinal Disease Surveillance using Social Media. 2016 Presented at: 2016 Digital Health Conference; April 11-13, 2016; Montreal, Canada.
64. Romano S, Martino SD, Kanhabua N, Mazzeo A, Nejdl W. Challenges in Detecting Epidemic Outbreaks from Social Networks. 2016 Presented at: 30th International Conference on Advanced Information Networking and Applications Workshops; Mar 23-25, 2016; Crans-Montana, Switzerland p. 978-971.
65. Grover S, Aujla GS. Twitter Data Based Prediction Model for Influenza Epidemic. 2015 Presented at: 2nd International Conference on Computing for Sustainable Global Development; March 11-13, 2015; New Delhi, India.
66. Talvis K, Chorianopoulos K, Kermanidis KL. Real-Time Monitoring of Flu Epidemics through Linguistic and Statistical Analysis of Twitter Messages. : IEEE; 2014 Presented at: 9th International Workshop on Semantic and Social Media Adaptation and Personalization; 2014; Corfu, Greece p. 83-87.
67. Culotta A. Detecting influenza outbreaks by analyzing Twitter messages. 2010 Presented at: Proceedings of the First Workshop on Social Media Analytics; July 25-28, 2010; Washington, DC, USA. [doi: 10.1145/1964858.1964874]
68. Zaldumbide J, Sinnott RO. Identification and Validation of Real-Time Health Events through Social Media. 2015 Presented at: Proceedings of the IEEE International Conference on Data Science and Data Intensive Systems; December ; Sydney, Australia. ISBN; December 11-13, 2015; Sydney, Australia.
69. Lampos V, Cristianini N. Tracking the flu pandemic by monitoring the social web. 2010 Presented at: 2nd International Workshop on Cognitive Information Processing; June 14-16, 2010; Elba, Italy. [doi: 10.1109/CIP.2010.5604088]
70. Lampos V, Cristianini N. Nowcasting events from the social web with statistical learning. ACM Trans Intell Syst Technol 2012 Sep 01;3(4):1-22. [doi: 10.1145/2337542.2337557]
71. Collier N, Doan S, Kawazoe A, Goodwin RM, Conway M, Tateno Y, et al. BioCaster: detecting public health rumors with a Web-based text mining system. Bioinformatics 2008 Dec 15;24(24):2940-2941 [FREE Full text] [doi: 10.1093/bioinformatics/btn534] [Medline: 18922806]
72. Chakraborty P, Khadivi P, Lewis B, Mahendiran A, Chen J, Butler P, et al. Forecasting a Moving Target: Ensemble Models for ILI Case Count Predictions. 2014 Presented at: SIAM International Conference on Data Mining; April 24-26, 2014; Pennsylvania, USA. [doi: 10.1137/1.9781611973440.30]
73. Nagar R, Yuan Q, Freifeld CC, Santillana M, Nojima A, Chunara R, et al. A case study of the New York City 2012-2013 influenza season with daily geocoded Twitter data from temporal and spatiotemporal perspectives. J Med Internet Res 2014;16(10):e236 [FREE Full text] [doi: 10.2196/jmir.3416] [Medline: 25331122]
74. Greaves F, Ramirez-Cano D, Millett C, Darzi A, Donaldson L. Harnessing the cloud of patient experience: using social media to detect poor quality healthcare. Br Med J Qual Saf 2013 Mar;22(3):251-255. [doi: 10.1136/bmjqs-2012-001527] [Medline: 23349387]
75. Bouazizi M, Otsuki Ohtsuki T. A pattern-based approach for sarcasm detection on Twitter. IEEE Access 2016;4:5477-5488. [doi: 10.1109/ACCESS.2016.2594194]

76. Pinheiro V, Pontes R, Furtado V. A #hashtagtokenizer for social media messages. Int J Comput Linguist Appl 2015;6(2):141-158 [FREE Full text]
77. Joshi A, Sharma V, Bhattacharyya P. Harnessing Context Incongruity for Sarcasm Detection. 2015 Presented at: 53rd Annual Meeting of the Association for Computational Linguistics; July 26-31, 2015; Beijing, China.
78. Signorini A, Segre AM, Polgreen PM. The use of Twitter to track levels of disease activity and public concern in the U.S. during the influenza A H1N1 pandemic. PLoS One 2011 May 04;6(5):e19467-e19410 [FREE Full text] [doi: 10.1371/journal.pone.0019467] [Medline: 21573238]

Abbreviations

AMT: Amazon Mechanical Turk
CDC: Center for Disease Control
CGD: consumer generated data
FSA: Food Standards Agency
IID: infectious intestinal disease
ILI: influenza-like illness
GP: general practitioner
k-NN: k-Nearest Neighbor
NB: Naïve Bayes Classifier
NLP: natural language processing
PHE: Public Health England
RSS: residual sum of squares
SVM: support vector machine

Edited by G Eysenbach; submitted 15.06.17; peer-reviewed by L Sbaffi, V Lampos; comments to author 22.09.17; revised version received 16.01.18; accepted 31.01.18; published 06.06.18.

Please cite as:
Oldroyd RA, Morris MA, Birkin M
Identifying Methods for Monitoring Foodborne Illness: Review of Existing Public Health Surveillance Techniques
JMIR Public Health Surveill 2018;4(2):e57
URL: http://publichealth.jmir.org/2018/2/e57/
doi:10.2196/publichealth.8218
PMID:

Review

Sentiment Analysis of Health Care Tweets: Review of the Methods Used

Sunir Gohil[1], BSc, MBBS; Sabine Vuik[1], MSc, PhD; Ara Darzi[1], MD, FRCS

[1]Imperial College London, Department of Surgery and Cancer, London, United Kingdom

Corresponding Author:
Sunir Gohil, BSc, MBBS
Imperial College London
Department of Surgery and Cancer
10th Floor QEQM Building, St Marys Hospital
Praed St, London
London, W2 1NY
United Kingdom
Phone: 44 7715552952
Fax: 44 1923854171
Email: sunir.gohil06@imperial.ac.uk

Abstract

Background: Twitter is a microblogging service where users can send and read short 140-character messages called “tweets.” There are several unstructured, free-text tweets relating to health care being shared on Twitter, which is becoming a popular area for health care research. Sentiment is a metric commonly used to investigate the positive or negative opinion within these messages. Exploring the methods used for sentiment analysis in Twitter health care research may allow us to better understand the options available for future research in this growing field.

Objective: The first objective of this study was to understand which tools would be available for sentiment analysis of Twitter health care research, by reviewing existing studies in this area and the methods they used. The second objective was to determine which method would work best in the health care settings, by analyzing how the methods were used to answer specific health care questions, their production, and how their accuracy was analyzed.

Methods: A review of the literature was conducted pertaining to Twitter and health care research, which used a quantitative method of sentiment analysis for the free-text messages (tweets). The study compared the types of tools used in each case and examined methods for tool production, tool training, and analysis of accuracy.

Results: A total of 12 papers studying the quantitative measurement of sentiment in the health care setting were found. More than half of these studies produced tools specifically for their research, 4 used open source tools available freely, and 2 used commercially available software. Moreover, 4 out of the 12 tools were trained using a smaller sample of the study’s final data. The sentiment method was trained against, on an average, 0.45% (2816/627,024) of the total sample data. One of the 12 papers commented on the analysis of accuracy of the tool used.

Conclusions: Multiple methods are used for sentiment analysis of tweets in the health care setting. These range from self-produced basic categorizations to more complex and expensive commercial software. The open source and commercial methods are developed on product reviews and generic social media messages. None of these methods have been extensively tested against a corpus of health care messages to check their accuracy. This study suggests that there is a need for an accurate and tested tool for sentiment analysis of tweets trained using a health care setting–specific corpus of manually annotated tweets first.

(JMIR Public Health Surveill 2018;4(2):e43) doi:10.2196/publichealth.5789

KEYWORDS

Twitter; social media

Introduction

Today's doctors and patients take to online platforms such as blogs, social media, and websites to convey opinions on health matters [1]. Infodemiology is "the science of distribution and determinants of information in an electronic medium, specifically the Internet, or in a population, with the ultimate aim to inform public health and public policy" [2]. Data can be collected and analyzed from social media such as Twitter in real time with the ability to survey public opinion (sentiment) toward a subject [3]. Bates and colleagues have described social media as a "perfect storm" in regard to patient-centered health care, which is a valuable source of data for the public and health organizations [4]. Twitter is one such place, being easy to use, cheap, and accessible. Twitter is a mobile microblogging and social networking service. There are currently 955 million registered Twitter users who can share messages that contain text, video, photos, or links to external sources. One-third of people with a social media profile use Twitter, with 75% accessing from a handheld device to convey an opinion [5,6].

Sentiment analysis allows the content of free-text natural language—that is, the words and symbols used in a message—to be examined for the intensity of positive and negative opinions and emotions. Sentiment analysis from social media is already a widely researched subject [7]. It is useful for business marketing to understand the public or consumer opinion toward their product [8]. Computerized software tools have been produced that automate the process of sentiment analysis, allowing large numbers of free-text comments to be processed into quantitative sentiment scores quickly, for example, positive or negative [7]. They are commonly based on text classifiers or machine learning processes. These tend to be commercially orientated, expensive, and focused on gathering opinion on a specific chosen product or service [9]. During the H1N1 outbreak, Chew et al conducted a content analysis of tweets [10]. In this study, they measured sentiment in a qualitative categorical way using content classifiers such as "humor" or "sarcasm." Accurate and automated sentiment analysis is challenging due to the subjectivity, complexity, and creativity of the language used [11].

Sentiment analysis in the health care setting is not a new phenomenon. Using only manual annotation of health care tweets, it has been found that 40% of messages contain some form of sentiment (either positive or negative) [12]. A manual method has also been used in the analysis of suicide notes and discharge summaries, where Cherry et al attempt to automate the manual process using machine learning approaches [13-15]. It was found that the manual classification of emotional text was difficult and inconsistent [13]. Greater positive sentiment within discharge summaries was associated with significantly decreased risk of readmission [14]. A study was also conducted measuring the sentiment of comments on the main National Health Service (NHS) website (NHS choices) over a 2-year period [16,17]. They found a strong agreement between the quantitative online ratings of health care providers and analysis of sentiment using their automated method.

Sentiment analysis has made its way into the mainstream analysis of Twitter-based health care research. Twitter is a popular platform as it allows data to be collected easily using their application programming interface. The limitations of other social media platforms such as Facebook are they do not allow such easy access to their data due to their varying privacy policies. It is not as easy to collect data in an open and automated way with other such media. The opinion of a tweet is found within the text portion of the tweet. This is captured in an unstructured, nonstandardized, free-text form. Accurately measuring the sentiment of a health care tweet represents an opportunity for understanding both the patient's and health care professional's opinion on a health subject [16]. Kent et al found that up to 40% of health care tweets contain some form of sentiment [12]. A validated tool for sentiment analysis of health care messages on Twitter would allow for the assessment of opinion on a mass scale [17]. Sentiment analysis in the medical setting offers a unique challenge as terms can have varying usage and meanings, and requires complementary context-specific features with a domain-specific lexicon [18]. The language used to convey sentiment in medicine is likely to be different than that toward a product, as the boundary between "patient," "consumer," and "customer" is difficult to define and terms can have varying usage and meanings [11,19]. Therefore, the sentiments may be expressed differently in a health care context [18].

To date, there has been no study looking at all the methods used for sentiment analysis on Twitter in the health care setting. Currently available sentiment analysis tools have not been developed based on a health care setting. SentiStrength [20], a popular open source software was based on nonspecific messages sent via MySpace [21]. Health care can be a very different environment based on many aspects. Being a public National Health Service [19], the boundary between "patient," "consumer," and "customer" is difficult to define in health care Therefore, currently available sentiment analysis methods may not be accurate.

The aim of this study was to review the methods used to measure sentiment for Twitter-based health care studies. The first objective was to review what methods of sentiment analysis have been used and in which health care setting. The second objective was to explore to what extent the methods were trained and validated for the study data, and if any justification for their methodology use was offered.

Methods

Identification and Screening

In May 2015, a computerized search of the literature was conducted, following Preferred Reporting Items for Systemic Reviews guidelines [22]. MEDLINE (OvidSP) and EMBASE (OvidSP) were searched using the terms. References were checked from papers and reviews, and citations were checked from included studies. The titles and abstracts were screened from the retrieved search to identify relevant studies. A supplementary hand search was carried out in September 2016 in key journals. Studies had to include one of the following search terms in the title, abstract, or keywords: "Twitter" or associated terms “tweet” or “microblog” and “Sentiment” or associated search terms “opinion” or “emoti” or “happi” or “Senti.” There were 3 inclusion criteria for the study. First, the study must have Twitter as its primary focus. The aim of this review was to explore research into the methods of sentiment analysis on Twitter messages only. Second, the papers must be relating to a health care subject. This included all aspects of health and health care delivery, health care research, policy, and organizational and professional use. Finally, papers that used a quantitative method to analyze both positive and negative sentiments of the messages, for example, “−1,” were included.

Eligibility and Inclusion

The studies were restricted to those published in English. A total of 69 full-text articles were assessed for eligibility. Of these, 15% (10/69) were rejected because they looked at social media in general (not Twitter specifically), for example, the use of social media by surgical colleagues [23]. Moreover, 36% (25/69) were rejected because the study did not pertain to health care, for example, public perceptions of nonmedical use of opioids [24]. Furthermore, 32% (22/69) papers were excluded because the sentiment analysis was either not measured, not quantitative or did not discuss positive and negative sentiments specifically, for example, characterizing sleep issues using Twitter [25]. The criteria used to compare the methods in each study looked at the method of tool production, in which setting it was used, and the method of testing the tool. For assessment, a comparison of the number of annotators used to manually annotate tweets, if any, and the level of agreement between them was used. Furthermore, the proportion of tweets used to train an algorithm compared with the final sample analyzed was also assessed.

Results

Overall Results

In total, 12 papers were found that satisfied all 3 inclusion criteria (see Table 1 for overview). These were published between 2011 and 2016 with data collected from Twitter between 2006 and 2016. Moreover, 2 papers examined global data, 9 in the United States, and 1 in the United Kingdom. Comments from 2 papers suggest that on an average 46% (92/2) of health care tweets contain some form of sentiment, that is, not neutral [12,26]. Many studies conducted analysis on public health–related subjects (n=7). In addition, 3 papers examined the sentiment toward an aspect of disease: the disease itself (n=1), symptoms (n=1), or treatment (n=1). Finally, 2 papers studied an emergency medical situation and a medical conference.

A total of 5 of the 12 studies conducted a manual sentiment analysis of a sample of their data using annotators to train their tool. One study used 13.58% (1000/7362) of their final data sample to train their developed method [34]. Three studies used an average of 0.7% of their total dataset to train their tool (1.46%, 250/17,098; 0.55%, 2216/404,065; and 0.1%, 250/198,499). One paper compared the accuracy of their chosen methods with a manually annotated corpus of their data [30]. Moreover, 2 papers from the group commented on justification of the sentiment analysis tools used.

There were 3 categories of sentiment analysis methods found (see Table 2), a tool specifically produced and trained for that study data, open source tools, and commercially available software. This distinction was made based on the required level of expertise in computer programming needed to implement that method and if predefined lexicons were used. Tools produced specifically for the study required the most amount of programming knowledge as these sometimes required the use of machine learning techniques to train a tool or rule-based methods. Alternatively, using commercially available software required the least knowledge as these are designed to be quick and easy to use. Half of the studies conducted quantitative sentiment analysis using an automated method developed by the study group themselves using algorithms or machine learning techniques. Moreover, 3 studies used commercially available sentiment analysis products. The remaining 3 papers used open source, freely available sentiment analysis software, which required little programming experience. In addition, 1 study from the open source and 1 from commercial method studies used a method of manual training to tailor the tool for their specific study data [33].

Table 1. Tools used for sentiment analysis.

Author	Year	Location	Subject area	Sentiment toward	Type of method
Bhattacharya et al [27]	2012	United States	Public health	25 Federal health agencies	Open source
Black et al [28]	2011	United States	Emergency medicine	2011 Japanese earthquake and tsunami	Commercial
Cole-Lewis et al [29]	2013-2014	United States	Public health	Electronic cigarette	Produced for study
Daniulaityte et al [30]	2016	United States	Public health	Sentiment toward drug-related tweets	Produced for study
Desai et al [31]	2011	United States	Medical conference	Twitter activity at Kidney Week 2011	Produced for study
Greaves et al [32]	2012	United Kingdom	Public health	Hospital quality	Commercial
Hawkins et al [33]	2015	United States	Public health	Hospital quality	Open source
Myslin et al [34]	2012	Global	Public health	Tobacco	Produced for study
Nwosu et al [35]	2015	Global	Disease specific	Palliative medicine	Open source
Ramagopalan et al [26]	2006-2014	United States	Disease treatment	Multiple sclerosis treatments	Open source
Sofean and Smith [36]	2013	United States	Public health	Tobacco	Produced for study
Tighe et al [37]	2015	United States	Disease symptoms	Pain	Produced for study

Table 2. Sentiment tools based on type of tool: KNN: k-nearest-neighbors; N/A: not applicable; NB: Naïve Bayes; SVM; support vector machines.

Author	Tool	Annotators	Kappa	Manually annotated sample	Sample size	Manually annotated compared with total sample, n (%)
Cole-Lewis et al [29]	Produced for study: machine learning classifiers based on 5 categories (NB, KNN, and SVM)	6	.64	250	17,098	250 (1.46)
Desai et al [31]	Produced for study: rule based using AFINN (Named after the author, Finn Arup Neilsen)	N/A	N/A	N/A	993	N/A
Daniulaityte et al [30]	Produced for study: logistic regression, NB, SVM	2	.68	3000	N/A	N/A
Myslin et al [34]	Produced for study: machine learning (NB, KNN, SVM)	2	>.7	1000	7362	1000 (13.58)
Sofean and Smith [36]	Produced for study: 5-fold validation using support vector machines (SVM's) model using Waikato Environment for Knowledge Analysis toolkit toolkit	N/A	N/A	500	N/A	N/A
Tighe et al [37]	Produced for study: rule based using AFINN	N/A	N/A	N/A	65,000	N/A
Bhattacharya et al [27]	Open source: SentiStrength	3	N/A	N/A	164,104	N/A
Hawkins et al [33]	Open source: machine learning classifier using Python library TextBlob	2+Amazon Mechanical Turk	>.79	2216	404,065	2216 (0.55)
Ramagopalan et al [26]	Open source: TwitteR R package + Jeffrey Breen's sentiment analysis code	N/A	N/A	N/A	60,037	N/A
Black et al [28]	Commercial: radian6	N/A	N/A	N/A	N/A	N/A
Greaves et al [32]	Commercial: TheySay	N/A	N/A	250	198,499	250 (0.13)
Nwosu et al [35]	Open source: TopsyPro	N/A	N/A	N/A	683,500	N/A

A total of 5 studies commented on the number of annotators used for the manual classification of sentiment to train their final tool (average=3 annotators, range 2-6). A single study used a method of outsourcing the task of manual classification to multiple anonymous annotators via Amazon Mechanical Turk [38].

Self-Produced Sentiment Analysis Tools

Of the 12 studies reviewed, 6 produced sentiment analysis tools within their own department, specifically designed for their study using already defined algorithms. Liu describes the different types of algorithms that can be used, and they produce different kinds of summaries [39,40]. Moreover, 2 different types of algorithms were found to be used, a standard supervised machine learning algorithm and a classification method (such as AFINN named after the author, Finn Arup Neilsen). These methods produce their own classifier trained to detect polarity using their original data. These may be different from the open source tools, which use already pretrained classifiers in premade software systems designed more toward an end user.

A total of 3 papers used a similar method of sentiment via categorization, all examining opinions toward smoking. Sofean et al produced an automated sentiment tool based on identifying 250 positive and 250 negative tweets from a smaller sample to train their tool [36]. There was no further detail into the annotation and analysis process. A limitation to their tool was that it screened out emoticons (symbols used to express emotion) before producing a tool. This is a method often used by users to convey emotion [39]. Myslin et al analyzed the sentiment toward emerging tobacco products on 7362 tweets, where Cole-Lewis et al looked specifically at sentiment toward electronic cigarettes on 17,098 tweets [29,34]. Neither of the studies commented on why a self-produced solution was used. Tweets were broadly categorized into “positive,” “neutral,” or “negative” by the annotators. The intensity of the sentiment was not recorded. To find the relationship between the sentiment and subject, 3 machine learning algorithms were used, Naïve Bayes, K-Nearest-Neighbor, and Support Vector Machine [41]. An automated sentiment analysis tool was produced based on the manual analysis of sentiment of a sample of tweets during the pilot phase of each study. This represented 13.58% (1000/7362) for Myslin. The study by Cole-Lewis used only 1.46% (250/17,098) of their total sample to train their algorithms. This represents a very small percentage of their sample and may result in their method being less accurate than intended. However, no comment is made by the study group to why only this number was used.

Desai et al used the AFINN (named after the author, Finn Arup Neilsen), to measure the sentiment of Twitter activity during Kidney Week 2011 from 993 tweets [31]. AFINN is a rule-based approach combined with statistical modeling to create a hybrid approach to sentiment classification [7]. This is based on comparing a sample of data with a list of weights of positive or negative keywords using the affective norms for English words dataset [42]. The AFINN consists of a list of manually labeled English words that have been given an integer value between −5 (highly negative) to +5 (highly positive). A value is assigned for each word in a tweet using the lexicon. The values are averaged to calculate the sentiment score for the whole message. This method has been validated for use in microblogs such as Twitter [43]. Tighe et al used this method to assess the sentiment of tweets pertaining to pain, suggesting a rule-based classifier has greater methodological advantage due to its deterministic results compared with human annotators which can have poor interannotator agreement with sentiment [37]. In addition, they supplemented AFINN with the use of emoticon terminology to enhance the accuracy of the rule-based classifier [39,44]. One study sought to compare different supervised machine learning (SML) techniques with each other, and to a rule-based open source lexicon for drug-related tweets [30]. They found that by using manually annotated tweets specifically from that subject to train SML techniques was more accurate than a preprepared lexicon due to the variation in language used. They also compare types of SML techniques to show that they all performed to a similar level.

Open Source Sentiment Software

Open source software is a computer software that has its source code made available to the public to modify [45]. The developers or copyright holders of the software give the rights to study and distribute the software for any purpose for free. Moreover, 4 papers used open source software for their sentiment analysis. None of these tools were initially produced using health care messages. Ramagopalan et al investigated the opinions of specific multiple sclerosis treatments using 60,037 tweets [26]. They used an open source sentiment analysis tool called package twitteR R [46] in combination with Jeffrey Breen’s sentiment analysis code [47]. This software was developed for the analysis of consumer sentiment toward a product and compares the frequency of positive or negative words against a predefined list. The overall sentiment score of each message is calculated by subtracting the number of negative words from the number of positive words. A sentiment score of >0 suggests that the message has an overall positive opinion. Of their dataset, 52% of messages contained a non-neutral sentiment. This study showed that there was a statistically significant difference in sentiment toward different types of multiple sclerosis medications. There was no comment on analysis of the tool itself or justification of its use.

Bhattacharya et al used SentiStrength [20,48], a popular open source software to analyze the sentiment of 164,104 tweets from 25 Federal Health Agencies in the United States and their 130 accounts. SentiStrength has been designed to measure the sentiment of short informal messages and has been widely used for Twitter analysis [49]. It was used in this case because it outperforms other lexical classifiers [42]. No manual sentiment analysis was conducted.

SentiStrength was developed in 2009 to extract sentiment strength from informal English text, giving a rating between −5 and +5. The algorithm was developed on an initial set of 2600 MySpace comments used for pilot testing. A set of 3 same gender (female) coders were used for initial testing and this was optimized by machine learning into its final version. It can detect positive emotion with 60.6% accuracy and negative emotion with 72.8% accuracy. SentiStrength outperforms a wide range of other machine learning approaches. SentiStrength has not yet been validated specifically for health care–based messages.

Hawkins et al measured patient-perceived quality of care in US hospitals using Twitter [33]. Over 404,000 tweets were analyzed for their sentiment and compared with established quality measures over a 1-year period. Natural language processing was used to measure the sentiment of the patient experience tweets. This was based on a Python library TextBlob [50]. TextBlob is trained from human annotated words commonly found in product reviews based on the Pattern Library [51]. The sentiment score can range from −1 to +1, with a score of 0 suggesting a tweet that is neutral. This was the first study that adopted Amazon Mechanical Turk [38] to use multiple outsourced anonymous curators to train their tool. They found a weak association between the positive sentiment toward a hospital and the readmission rate.

Commercial Software

There are numerous commercial software packages available to analyze the sentiment of tweets. These range in price depending on the number of tweets or duration of use. In this study, 2 papers were found using commercial software. Neither tool was developed with health care messages as its foundation, and no justification for their use is offered for either.

The largest number of messages analyzed by Nwosu measured the sentiment of over 683,000 tweets based around palliative medicine and end of life care [35]. Discussion about end of life can be difficult and sometimes missed [52]. TopsyPro was used to measure the sentiment of tweets [53]. This software was created in 2015 as an Web based tool for Twitter analytics and sentiment analysis and is based on an annual subscription costing US $12,000 per year per named user (for the “Pro” version which enables more detailed analysis). There is no information currently available on the methods used by Topsy Labs, Inc. on how the sentiment analysis is conducted.

Radian6 [54] is another piece of “listening” social media software to collect and analyze data. It has been previously used to collect data during a medical conference, with analysis focused on the major Twitter influencers [55]. The software does not require the user to have any programming knowledge and is deigned to be easy to use. Black et al used this software to analyze tweets based around public health emergency response during the Japanese earthquake and tsunami in March 2011. There was no comment on why this software was used. Radian6 can “listen” automatically to large-scale Twitter conversation based on specific keywords.

A study conducted by Greaves et al was found looking at hospital quality in the United Kingdom, and it measured the sentiment of over 198,000 tweets directed toward NHS hospitals in 2012 [32]. The commercially available software used was developed by TheySay Ltd (Oxford, UK). TheySay is based on compositional sentiment parsing, described by work from Moilanen and Pulman, using 5 automated ways of natural language processing [56]. For academic purposes, the software costs roughly £350 for a similar volume of data to the mentioned study to be analyzed.

Discussion

Principal Findings

On average, 46% (92/2) of health-based tweets contain some form of positive or negative sentiment [12,26]. A relationship between sentiment on Twitter and hospital statistics has already been proven [33]. It is important to conduct sentiment analysis for health care tweets that is accurate and consistent. This study has found that there is a large disparity in the types of methods used, from basic categorizations to seemingly sophisticated and expensive commercially available software. Between the same subject matter such as hospital quality, different sentiment analysis methods have been used which makes it difficult to compare the results between the two [32,33]. Chew et al conducted a content analysis of tweets during the 2009 H1N1 outbreak and chose to use only a qualitative method for sentiment analysis of tweets, categorizing tweets based on emotive words, for example, “Humour” or “Concern” [10]. On the basis of complexity of implementation, 3 broad categories of methods have emerged: (1) self-produced methods using algorithms, (2) open source methods, and (3) commercially available software. Only 1 method in this study was produced with health care language as its foundation using a corpus of manually annotated health care setting–specific tweets for training [30]. Many methods were based on tools trained on product reviews and nonspecific social media messages that may not be appropriate for use in the health care setting [20,57]. The language used to convey sentiment in medicine is likely to be different than that toward a product as the boundary between “patient,” “consumer,” and “customer” is difficult to define and terms can have varying usage and meanings [11,18,19]. Health-related tweets represent a unique type of content, and their communication on Twitter carries special characteristics as found in pain-related tweets [37].

Most studies did not justify the reason for their selected method. Furthermore, there was no evidence of analysis of accuracy of the method before being used for the larger respective data. Researchers tend to assume a method selected will be accurate. Most self-produced methods train their tool using a very small percentage of their final dataset, in one case less than 2% [29]. A formal process for checking the accuracy occurred in one of the author’s study that compared types of supervised machine learning techniques. Software products and open source tools being currently used tend to be designed originally to identify opinions about products in the commercial setting rather than behaviors. This questions their accuracy when used in a medical setting.

Recommendations

This research shows that different approaches are used for the sentiment analysis of tweets in the health care setting. The evidence suggests that there is a need for the production and analysis of accuracy of a sentiment analysis tool trained using setting-specific health care tweets. Twitter is used globally, and health care can vary greatly depending on the setting. On the basis of this study, such a tool would ideally be trained using a health care subject-specific corpus of labeled tweets to train supervised machine learning classifiers [30]. Semantic

Evaluation Exercises (SemEval 2016) held in San Diego is an event where programmers are tasked with producing a sentiment analysis tool on a range of Twitter subjects such as a political candidate or product, using a pre-annotated corpus. This collaborative approach could be used to produce a more advanced and accurate tool for the health care setting using subject-specific lexicons and complementary health care–based features [11,18,58]. Furthermore, it could measure the intensity of sentiment using an aggregation of methods (eg, emoticons, natural language processing, and supervised machine learning), and it could check for accuracy against a slightly larger manually annotated dataset before being used on much larger sample sizes. This could allow future research in health care–based tweets to accurately and consistently measure the sentiment of setting specific health care–based messages.

Conflicts of Interest

None declared.

References

1. Afyouni S, Fetit AE, Arvanitis TN. #DigitalHealth: exploring users' perspectives through social media analysis. Stud Health Technol Inform 2015;213:243-246. [Medline: 26153005]
2. Eysenbach G. Infodemiology and infoveillance: framework for an emerging set of public health informatics methods to analyze search, communication and publication behavior on the Internet. J Med Internet Res 2009 Mar 27;11(1):e11 [FREE Full text] [doi: 10.2196/jmir.1157] [Medline: 19329408]
3. Eysenbach G. Infodemiology and infoveillance tracking online health information and cyberbehavior for public health. Am J Prev Med 2011 May;40(5 Suppl 2):S154-S158. [doi: 10.1016/j.amepre.2011.02.006] [Medline: 21521589]
4. Rozenblum R, Bates D. Patient-centred healthcare, social media and the internet: the perfect storm? BMJ Qual Saf 2013 Feb 01;22(3):183-186. [doi: 10.1136/bmjqs-2012-001744]
5. Ofcom. 2015. The communications market report URL: http://stakeholders.ofcom.org.uk/binaries/research/cmr/cmr15/icmr15/icmr_2015.pdf [WebCite Cache ID 6gjgoy6Ex]
6. Lunden I. Techcrunch. 2013. Mobile twitterm+ (75%) access from handheld devices monthly, 65% of ad sales come from mobile URL: http://techcrunch.com/2013/10/03/mobile-twitter-161m-access-from-handheld-devices-each-month-65-of-ad-revenues-coming-from-mobile/ [accessed 2016-04-13] [WebCite Cache ID 6gjgwD7Lv]
7. Pang B, Lee L. Opinion mining and sentiment analysis. Foundations and trends in information retrieval 2008 Jul;2(1-2):1-35.
8. Liu B, Zhang L. A survey of opinion mining and sentiment analysis. Mining Text Data 2012:415-463. [doi: 10.1007/978-1-4614-3223-4]
9. Nasukawa T. Sentiment analysis: capturing favorability using natural language processing. 2003 Jan 01 Presented at: Proceedings of the 2nd International Conference on Knowledge; October 23-25, 2003; Sanibel Island, FL, USA. [doi: 10.1145/945645.945658]
10. Chew C, Eysenbach G. Pandemics in the age of twitter: content analysis of Tweets during the 2009 H1N1 outbreak. PLoS One 2010 Nov 29;5(11):e14118 [FREE Full text] [doi: 10.1371/journal.pone.0014118] [Medline: 21124761]
11. Mohammad S. 9 – Sentiment analysis: detecting valence, emotions, and other affectual states from text. Emotion Measurement 2016:201-237. [doi: 10.1016/B978-0-08-100508-8.00009-6]
12. Kent E, Prestin A, Gaysynsky A, Galica K, Rinker R, Graff K, et al. “Obesity is the new major cause of cancer”: connections between obesity and cancer on facebook and twitter. J Canc Educ 2015 Apr 14;31(3):453-459. [doi: 10.1007/s13187-015-0824-1]
13. Danno K, Horio T. Sunburn cell: factors involved in its formation. Photochem Photobiol 1987 May;45(5):683-690. [Medline: 3299408]
14. Treves AJ, Carnaud C, Trainin N, Feldman M, Cohen IR. Enhancing T lymphocytes from tumor-bearing mice suppress host resistance to a syngeneic tumor. Eur J Immunol 1974 Nov;4(11):722-727. [doi: 10.1002/eji.1830041104] [Medline: 4547711]
15. Yadav RN. Isocitrate dehydrogenase activity and its regulation by estradiol in tissues of rats of various ages. Cell Biochem Funct 1988 Jul;6(3):197-202. [doi: 10.1002/cbf.290060308] [Medline: 3409480]
16. Greaves F, Ramirez-Cano D, Millett C, Darzi A, Donaldson L. Use of sentiment analysis for capturing patient experience from free-text comments posted online. J Med Internet Res 2013 Nov 01;15(11):e239-e251 [FREE Full text] [doi: 10.2196/jmir.2721] [Medline: 24184993]
17. Alemi F, Torii M, Clementz L, Aron DC. Feasibility of real-time satisfaction surveys through automated analysis of patients' unstructured comments and sentiments. Qual Manag Health Care 2012;21(1):9-19. [doi: 10.1097/QMH.0b013e3182417fc4]
18. Denecke K, Deng Y. Sentiment analysis in medical settings: new opportunities and challenges. Artif Intell Med 2015 May;64(1):17-27. [doi: 10.1016/j.artmed.2015.03.006]

19. Alemi F, Torii M, Clementz L, Aron DC. Feasibility of real-time satisfaction surveys through automated analysis of patients' unstructured comments and sentiments. Qual Manag Health Care 2012;21(1):9-19. [doi: 10.1097/QMH.0b013e3182417fc4] [Medline: 22207014]
20. Thelwall M, Buckley K, Paltoglou G, Cai D, Kappas A. Sentiment strength detection in short informal text. J Am Soc Inf Sci 2010 Dec 15;61(12):2544-2558. [doi: 10.1002/asi.21416]
21. MySpace. 2016. Featured content on myspace URL: https://myspace.com/ [WebCite Cache ID 6gDjWFfGw]
22. Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. Ann Intern Med 2009 Aug 18;151(4):264-9, W64. [Medline: 19622511]
23. Loeb S, Bayne CE, Frey C, Davies BJ, Averch TD, Woo HH, American Urological Association Social Media Work Group. Use of social media in urology: data from the American Urological Association (AUA). BJU Int 2014 Jun;113(6):993-998 [FREE Full text] [doi: 10.1111/bju.12586] [Medline: 24274744]
24. Chan B, Lopez A, Sarkar U. The canary in the coal mine tweets: social media reveals public perceptions of non-medical use of opioids. PLoS One 2015;10(8):e0135072 [FREE Full text] [doi: 10.1371/journal.pone.0135072] [Medline: 26252774]
25. de Chudnovsky A W. [The contemporary face and the prediction of growth]. Trib Odontol (B Aires) 1973;57(10):294-6 passim. [Medline: 4526927]
26. Adelman AG, Wigle ED, Ranganathan N, Webb GD, Kidd BS, Bigelow WG, et al. The clinical course in muscular subaortic stenosis. A retrospective and prospective study of 60 hemodynamically proved cases. Ann Intern Med 1972 Oct;77(4):515-525. [Medline: 4264640]
27. Bhattacharya S, Srinivasan P, Polgreen P. Engagement with health agencies on twitter. PLoS One 2014;9(11):e112235 [FREE Full text] [doi: 10.1371/journal.pone.0112235] [Medline: 25379727]
28. Silen W, Machen TE, Forte JG. Acid-base balance in amphibian gastric mucosa. Am J Physiol 1975 Sep;229(3):721-730. [doi: 10.1152/ajplegacy.1975.229.3.721] [Medline: 2015]
29. Cole-Lewis H, Varghese A, Sanders A, Schwarz M, Pugatch J, Augustson E. Assessing electronic cigarette-related tweets for sentiment and content using supervised machine learning. J Med Internet Res 2015 Aug 25;17(8):e208 [FREE Full text] [doi: 10.2196/jmir.4392] [Medline: 26307512]
30. Daniulaityte R, Chen L, Lamy FR, Carlson RG, Thirunarayan K, Sheth A. “When 'bad' is 'good'”: identifying personal communication and sentiment in drug-related tweets. JMIR Public Health Surveill 2016 Oct 24;2(2):e162 [FREE Full text] [doi: 10.2196/publichealth.6327] [Medline: 27777215]
31. Desai T, Shariff A, Shariff A, Kats M, Fang X, Christiano C, et al. Tweeting the meeting: an in-depth analysis of twitter activity at kidney week 2011. PLoS One 2012;7(7):e40253 [FREE Full text] [doi: 10.1371/journal.pone.0040253] [Medline: 22792254]
32. Greaves F, Laverty AA, Cano DR, Moilanen K, Pulman S, Darzi A, et al. Tweets about hospital quality: a mixed methods study. BMJ Qual Saf 2014 Oct;23(10):838-846 [FREE Full text] [doi: 10.1136/bmjqs-2014-002875] [Medline: 24748372]
33. Hawkins J, Brownstein JS, Tuli G, Runels T, Broecker K, Nsoesie EO, et al. Measuring patient-perceived quality of care in US hospitals using Twitter. BMJ Qual Saf 2016 Dec;25(6):404-413 [FREE Full text] [doi: 10.1136/bmjqs-2015-004309] [Medline: 26464518]
34. Myslín M, Zhu SH, Chapman W, Conway M. Using twitter to examine smoking behavior and perceptions of emerging tobacco products. J Med Internet Res 2013 Aug 29;15(8):e174 [FREE Full text] [doi: 10.2196/jmir.2534] [Medline: 23989137]
35. Nwosu A, Debattista M, Rooney C, Mason S. Social media and palliative medicine: a retrospective 2-year analysis of global twitter data to evaluate the use of technology to communicate about issues at the end of life. BMJ Support Palliat Care 2015 Jun;5(2):207-212. [doi: 10.1136/bmjspcare-2014-000701] [Medline: 25183713]
36. Sofean M, Smith M. Sentiment analysis on smoking in social networks. Stud Health Technol Inform 2013;192:1118. [Medline: 23920892]
37. Tighe PJ, Goldsmith RC, Gravenstein M, Bernard HR, Fillingim RB. The painful tweet: text, sentiment, and community structure analyses of tweets pertaining to pain. J Med Internet Res 2015 Apr 02;17(4):e84 [FREE Full text] [doi: 10.2196/jmir.3769] [Medline: 25843553]
38. Amazon. Amazon. 2015. Amazon mechanical turk - welcome URL: https://www.mturk.com/ [WebCite Cache ID 6gDjkM7J3]
39. Liu K, Li W, Guo M. Emoticon smoothed language models for twitter sentiment analysis. 2012 Presented at: AAAI'12 Proceedings of the Twenty-Sixth AAAI Conference on Artificial Intelligence; July 22-26, 2012; Toronto, Ontario, Canada p. 1678-1684.
40. Liu S, Yamada M, Collier N, Sugiyama M. Change-point detection in time-series data by relative density-ratio estimation. Neural Netw 2013 Jul;43:72-83. [doi: 10.1016/j.neunet.2013.01.012] [Medline: 23500502]
41. Franklin J. The elements of statistical learning: data mining, inference and prediction. The Mathematical Intelligencer 2008 Nov 12;27(2):83-85. [doi: 10.1007/BF02985802]
42. Nielsen F. arxiv. 2011. A new ANEW: evaluation of a word list for sentiment analysis in microblogs URL: https://arxiv.org/abs/1103.2903 [WebCite Cache ID 6yE38xA6F]
43. Kouloumpis E. Twitter sentiment analysis: the good the bad and the OMG!. 2011 Presented at: Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media; July 17-21, 2011; Barcelona, Catalonia, Spain p. 538-541.

44. Davidov D, Tsur O, Rappoport A. Enhanced sentiment learning using Twitter hashtags and smileys. 2010 Presented at: COLING '10 Proceedings of the 23rd International Conference on Computational Linguistics: Posters; August 23-27, 2010; Beijing, China.
45. Opensource. 2015. What is open source? URL: https://opensource.com/resources/what-open-source [WebCite Cache ID 6gjjUVOPo]
46. Soranaka K, Matsushita S. Relationship between emotional words and emoticons in tweets. 2012 Presented at: 2012 Conference on Technologies and Applications of Artificial Intelligence (TAAI); November 16-18, 2012; Tainan, Taiwan. [doi: 10.1109/TAAI.2012.30]
47. Wordpress. 2011. slides from my R tutorial on Twitter text mining #rstats URL: https://jeffreybreen.wordpress.com/2011/07/04/twitter-text-mining-r-slides/ [accessed 2016-04-13] [WebCite Cache ID 6gjqcREzI]
48. Sentistrength. 2015. SentiStrength - sentiment strength detection in short texts - sentiment analysis, opinion mining URL: http://sentistrength.wlv.ac.uk/ [WebCite Cache ID 6gDjqyjzI]
49. Thelwall M, Buckley K, Paltoglou G. Sentiment in twitter events. J Am Soc Inf Sci 2010 Dec 06;62(2):406-418. [doi: 10.1002/asi.21462]
50. Loria S. TextBlob. TextBlob: simplified text processing URL: https://textblob.readthedocs.org/en/dev/ [accessed 2016-04-13] [WebCite Cache ID 6gjqofstl]
51. Computational Linguistics & Psycholinguistics Research Center (CLiPS). 2015. pattern.en URL: http://www.clips.ua.ac.be/pages/pattern-en [WebCite Cache ID 6gDjvzIX2]
52. Mori M, Shimizu C, Ogawa A, Okusaka T, Yoshida S, Morita T. A national survey to systematically identify factors associated with oncologists' attitudes toward end-of-life discussions: what determines timing of end-of-life discussions? Oncologist 2015 Nov;20(11):1304-1311 [FREE Full text] [doi: 10.1634/theoncologist.2015-0147] [Medline: 26446232]
53. Hortanoticias. 2015. Twitter analytics by Topsy. Search hashtags, sentiment & trends URL: https://www.hortanoticias.com/wp-content/uploads/2015/03/Twitter-Analytics-by-Topsy-HortaTuits.pdf [accessed 2016-03-23] [WebCite Cache ID 6gDk08ji2]
54. Salesforce. 2015. Social studio migration URL: https://www.salesforce.com/form/marketingcloud/social-studio-migration.jsp [accessed 2016-04-13] [WebCite Cache ID 6gjdAk1IS]
55. Ferguson C, Inglis SC, Newton PJ, Cripps PJ, Macdonald PS, Davidson PM. Social media: A tool to spread information: a case study analysis of twitter conversation at the cardiac society of Australia & New Zealand 61st annual scientific meeting 2013. Collegian 2014 Jun;21(2):89-93. [doi: 10.1016/j.colegn.2014.03.002] [Medline: 25109206]
56. Moilanen K, Pulman S. Sentiment composition. 2007 Presented at: Proceedings of Recent Advances in Natural Language Processing (RANLP 2007); September 27-29, 2007; Borovets, Bulgaria p. 378-382.
57. Hawkins C, Duszak R, Rawson JV. Social media in radiology: early trends in twitter microblogging at radiology's largest international meeting. J Am Coll Radiol 2014 Apr;11(4):387-390. [doi: 10.1016/j.jacr.2013.07.015] [Medline: 24139963]
58. Mohammad S, Turney P. Crowdsourcing a word-emotion association lexicon. Comput Intell 2013 Aug;29(3):436-465. [doi: 10.1111/j.1467-8640.2012.00460.x]

Abbreviations

KNN: k-nearest-neighbors
NB: Naïve Bayes
NHS: National Health Service
SML: Supervised Machine Learning
SVM: support vector machines

Edited by G Eysenbach; submitted 18.04.16; peer-reviewed by S Mohammad, R Hilscher, T Hernández, M Larsen, H Singh; comments to author 25.07.16; revised version received 31.10.16; accepted 14.03.17; published 23.04.18.

Please cite as:
Gohil S, Vuik S, Darzi A
Sentiment Analysis of Health Care Tweets: Review of the Methods Used
JMIR Public Health Surveill 2018;4(2):e43
URL: http://publichealth.jmir.org/2018/2/e43/
doi:10.2196/publichealth.5789
PMID:29685871

Original Paper

Public Awareness of Uterine Power Morcellation Through US Food and Drug Administration Communications: Analysis of Google Trends Search Term Patterns

Lauren N Wood[1], MD; Juzar Jamnagerwalla[2], MD; Melissa A Markowitz[3], BA; D Joseph Thum[2], MD; Philip McCarty[3], BA; Andrew R Medendorp[1], MD; Shlomo Raz[1], MD; Ja-Hong Kim[1], MD

[1]Department of Urology, University of California, Los Angeles, Los Angeles, CA, United States
[2]Division of Urology, Cedars-Sinai Medical Center, Los Angeles, CA, United States
[3]University of California, Los Angeles, Los Angeles, CA, United States

Corresponding Author:
Lauren N Wood, MD
Department of Urology
University of California, Los Angeles
200 Medical Plaza, Suite 140
Los Angeles, CA,
United States
Phone: 1 310 794 0206
Email: laurennwood@gmail.com

Abstract

Background: Uterine power morcellation, where the uterus is shred into smaller pieces, is a widely used technique for removal of uterine specimens in patients undergoing minimally invasive abdominal hysterectomy or myomectomy. Complications related to power morcellation of uterine specimens led to US Food and Drug Administration (FDA) communications in 2014 ultimately recommending against the use of power morcellation for women undergoing minimally invasive hysterectomy. Subsequently, practitioners drastically decreased the use of morcellation.

Objective: We aimed to determine the effect of increased patient awareness on the decrease in use of the morcellator. Google Trends is a public tool that provides data on temporal patterns of search terms, and we correlated this data with the timing of the FDA communication.

Methods: Weekly relative search volume (RSV) was obtained from Google Trends using the term “morcellation.” Higher RSV corresponds to increases in weekly search volume. Search volumes were divided into 3 groups: the 2 years prior to the FDA communication, a 1-year period following, and thereafter, with the distribution of the weekly RSV over the 3 periods tested using 1-way analysis of variance. Additionally, we analyzed the total number of websites containing the term “morcellation” over this time.

Results: The mean RSV prior to the FDA communication was 12.0 (SD 15.8), with the RSV being 60.3 (SD 24.7) in the 1-year after and 19.3 (SD 5.2) thereafter (P<.001). The mean number of webpages containing the term “morcellation” in 2011 was 10,800, rising to 18,800 during 2014 and 36,200 in 2017.

Conclusions: Google search activity about morcellation of uterine specimens increased significantly after the FDA communications. This trend indicates an increased public awareness regarding morcellation and its complications. More extensive preoperative counseling and alteration of surgical technique and clinician practice may be necessary.

(JMIR Public Health Surveill 2018;4(2):e47) doi:10.2196/publichealth.9913

KEYWORDS

Google; internet search activity; FDA safety communication; uterine morcellation

Introduction

Uterine power morcellation, where the uterus is shred into smaller pieces, is a widely used technique for removal of uterine specimens in patients undergoing minimally invasive abdominal hysterectomy or myomectomy [1]. The power morcellator has also been used by many practitioners during hysterectomy for women undergoing concomitant prolapse repair [2]. Morcellation was a seemingly attractive option to minimize the size of incision needed to remove a uterine specimen and decrease postoperative pain, length of hospital stay, and potential risk of hernia.

In 2014, the US Food and Drug Administration (FDA) published two safety communications regarding uterine power morcellation. The first communication, released on April 17, 2014, specifically discouraged the use of laparoscopic power morcellation during hysterectomy or myomectomy of uterine fibroids due to a small risk of spreading undiagnosed uterine sarcoma, despite the low risk of finding an unsuspected uterine malignancy [3]. On November 24, 2014, a second FDA communication was released, this time warning against the use of laparoscopic power morcellation in the majority of women undergoing treatment of fibroids with either a myomectomy or hysterectomy, again citing the low risk of spread of undiagnosed uterine sarcoma as the rationale [4]. Furthermore, practitioners were urged to communicate this information directly with patients. Subsequently, the use of the morcellator drastically decreased [5].

Previous studies have shown that public awareness of health-related spectacles can be demonstrated using Google Trends, a free publicly available tool that provides data on temporal patterns of search terms [6,7]. Google Trends can be used for causal inference, description, or surveillance of various health-related research topics [6]. We hypothesized that Google Trends may be used to determine public interest in uterine power morcellation by correlating trends with the timing of the FDA warning.

Methods

Google Trends is a free publicly available tool that provides data on the number of times a certain term is searched over time on the internet search engine Google. We used it to examine public awareness regarding uterine power morcellation. Data are normalized to a reference population and then scaled to create a weekly relative search volume (RSV) ranging from 0 to 100, with the highest search activity scored at 100 and search activity at all other times scored relative to that peak [7]. A higher RSV represents an increase in search volume compared to other time frames, with an RSV of 100 representing a maximum search volume over a given temporal period.

We performed a Google Trends search using the term “uterine morcellation” to obtain RSVs averaged over 7-day periods. We then compared the weekly RSVs before the initial FDA communication in 2014, 1 year after the FDA warning, and thereafter. Analysis of variance test was used to explore the relationship between the RSVs during these time periods.

Additionally, various internet search engines provide information about the absolute number of webpage results available for a search term reported annually, which can provide insight into availability of information to the public. An increase in the number of webpages containing a search term year to year indicates an increased public interest in that term; thus, we examined the annual number of websites containing “morcellation” using the search engine Bing for a 3-year period from 2011 to 2014 (the 3 years prior to the communications), 2014 (the year of the communications), and 2014 to 2017 (the 3 years following the communication). All statistical analyses were performed using Stata version 13.1 (StataCorp LLC).

Results

The mean RSV prior to the initial 2014 FDA safety communication was 12, indicating very little baseline search volume for uterine morcellation (Table 1). This significantly increased to an average of 60.3 in the year following the FDA safety communication and decreased to 19.3 thereafter (*P*<.001). A peak RSV of 100 occurred twice, once in June 2014, between the 2 communications, and again in December 2014, after the second FDA safety communication (Figure 1).

Table 1. Weekly relative search volume for the term “uterine morcellation.”

Time period	Relative search volume, mean (SD)	*P* value
Prior to US Food and Drug Administration communication	12.0 (15.8)	<.001
One year after communication	60.3 (24.7)	<.001
Thereafter	19.3 (5.2)	<.001

Figure 1. Weekly relative search volume from Google Trends for the term "morcellation."

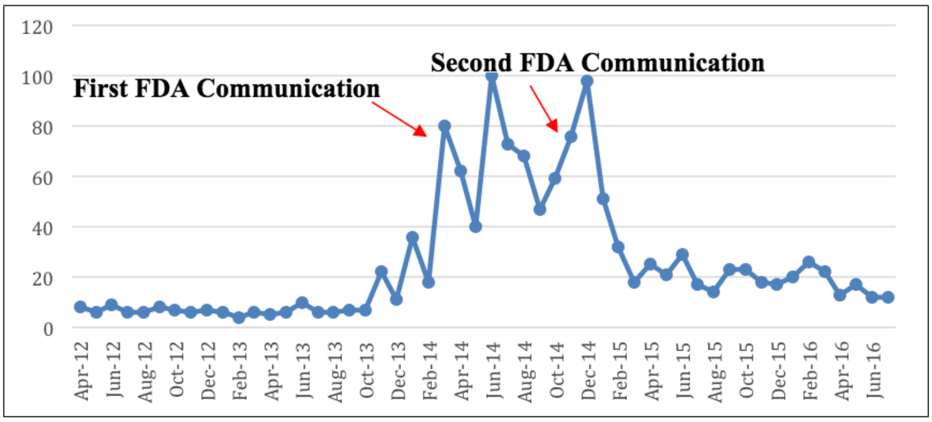
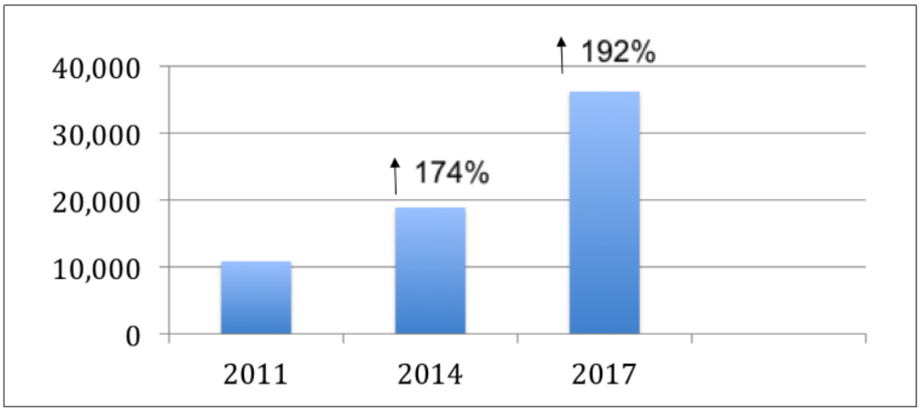


Table 2. Number of webpages containing the term “morcellation.”

Time period	Number of webpages
2011	10,800
2014	18,880
2017	36,200

Figure 2. Webpages containing the term "morcellation."



The mean number of webpages containing the term “morcellation” in 2011 was 10,800 (Table 2). This rose significantly to 18,800 during the year of the communication in 2014, representing an increase of 174% (Figure 2). The largest increase in the annual number of webpages was seen in 2017, when the number rose to 36,200, representing an increase of 192%.

Discussion

Principal Findings

Uterine fibroids are common, with up to 80% of women having fibroids by age 50 years [8]. The FDA statements to discourage the use of power morcellation were driven by the risk of undiagnosed uterine malignancy. Ackenbom et al [9] demonstrated that the rate of occult malignancy in patients undergoing hysterectomy for pelvic organ prolapse was significantly lower than in other patient populations. In addition to the risk of cancer progression, morcellation may increase the risk of parasitic leiomyomata and iatrogenic endometriosis [10].

The alternatives to power morcellation include open surgery, extending a laparoscopic incision to remove the specimen, or in-bag morcellation. The use of a bag for morcellation has been shown to increase operative time [11]. Given the low risk of malignancy, it would be difficult to demonstrate that in-bag morcellation would result in a decrease in cancer risk, and no evidence for this is currently available.

As technology continues to evolve, patient awareness of publicly available health information becomes increasingly important for physicians to consider when counseling patients. The use of Google Trends has become increasingly important in health care, with a systematic review of over 70 papers showing a 7-fold increase in the number of Google Trends publications from 2009 to 2013 [6]. Matta et al [12] demonstrated that urologists are drastically increasing their use of technology and social media, and patients are likely to seek out health care information both from their physicians and online. Previous authors have used Google Trends to demonstrate that patients in the United States are more likely than those in several other countries to search for topics related to cancer [13] such as uterine power morcellation.

Previous authors have shown an increase in the search term “pelvic organ prolapse” using Google Trends that was associated with the 2011 FDA safety communication [14]. Similarly, we used Google Trends to assess the impact of the 2014 FDA safety communications regarding the use of the power morcellator. This decrease in use may be at least partially driven by patient awareness, as shown by substantial increases in search volume and total number of webpages containing morcellation around the time of the FDA safety communications.

Limitations

Our study does have limitations. By limiting our analysis to Google Trends and Bing, we do not capture any relevant volume from other search engines. However, Google is the most widely used search engine in the United States. Furthermore, internet users tend to be younger, while patients undergoing hysterectomy are likely to be older. However, use of the internet for medical research is gaining popularity among older patients, with nearly 75% of primary care patients over the age of 65 years using the internet and nearly half using it to access health information [15]. Additionally, decreased use of the morcellator could result from provider-driven decisions and not just patient awareness.

Conclusion

In summary, the decrease in the use of the power morcellator for uterine specimens at the time of hysterectomy that followed the 2014 FDA safety communications may be related to a significant increase in Google search volume and mean number of webpages containing the term “morcellation.” This suggests that patient awareness may be in part driving the decreased use of uterine power morcellation and may indicate the need for clinicians to expand their scope of preoperative counseling or alter practice patterns and surgical technique.

Conflicts of Interest

None declared.

References

1. Perkins RB, Handal-Orefice R, Hanchate AD, Lin M, Paasche-Orlow MK. Risk of undetected cancer at the time of laparoscopic supracervical hysterectomy and laparoscopic myomectomy: implications for the use of power morcellation. Womens Health Issues 2016;26(1):21-26. [doi: 10.1016/j.whi.2015.09.008] [Medline: 26701205]
2. Vallabh-Patel V, Saiz C, Salamon C, Francis A, Pagnillo J, Culligan P. Prevalence of occult malignancy within morcellated specimens removed during laparoscopic sacrocolpopexy. Female Pelvic Med Reconstr Surg 2016;22(4):190-193. [doi: 10.1097/SPV.0000000000000257] [Medline: 26945272]
3. Laparoscopic uterine power morcellation in hysterectomy and myomectomy: FDA Safety Communication. Silver Spring: US Food and Drug Administration; 2014 Apr 17. URL: https://wayback.archive-it.org/7993/20170722215731/https://www.fda.gov/MedicalDevices/Safety/AlertsandNotices/ucm393576.htm [accessed 2018-04-19] [WebCite Cache ID 6ynlfuv7I]
4. UPDATED—Laparoscopic uterine power morcellation in hysterectomy and myomectomy: FDA Safety Communication. Silver Spring: US Food and Drug Administration; 2014 Nov 24. URL: https://wayback.archive-it.org/7993/20170722215727/https://www.fda.gov/MedicalDevices/Safety/AlertsandNotices/ucm424443.htm [accessed 2018-04-19] [WebCite Cache ID 6ynlv3RYu]
5. Wright JD, Chen L, Burke WM, Hou JY, Tergas AI, Ananth CV, et al. Trends in use and outcomes of women undergoing hysterectomy with electric power morcellation. JAMA 2016;316(8):877-878. [doi: 10.1001/jama.2016.9432] [Medline: 27552622]
6. Nuti SV, Wayda B, Ranasinghe I, Wang S, Dreyer RP, Chen SI, et al. The use of Google Trends in health care research: a systematic review. PLoS One 2014;9(10):e109583 [FREE Full text] [doi: 10.1371/journal.pone.0109583] [Medline: 25337815]
7. Google Trends. URL: https://trends.google.com/trends/yis/2016/US/ [accessed 2018-04-19] [WebCite Cache ID 6ynm0hofr]
8. Nieboer TE, Johnson N, Lethaby A, Tavender E, Curr E, Garry R, et al. Surgical approach to hysterectomy for benign gynaecological disease. Cochrane Database Syst Rev 2009 Jul 08(3):CD003677. [doi: 10.1002/14651858.CD003677.pub4] [Medline: 19588344]
9. Ackenbom MF, Giugale LE, Wang Y, Shepherd JP. Incidence of occult uterine pathology in women undergoing hysterectomy with pelvic organ prolapse repair. Female Pelvic Med Reconstr Surg 2016;22(5):332-335. [doi: 10.1097/SPV.0000000000000283] [Medline: 27171317]
10. Pereira N, Buchanan TR, Wishall KM, Kim SH, Grias I, Richard SD, et al. Electric morcellation-related reoperations after laparoscopic myomectomy and nonmyomectomy procedures. J Minim Invasive Gynecol 2015 Feb;22(2):163-176. [doi: 10.1016/j.jmig.2014.09.006] [Medline: 25218993]
11. Vargas MV, Cohen SL, Fuchs-Weizman N, Wang KC, Manoucheri E, Vitonis AF, et al. Open power morcellation versus contained power morcellation within an insufflated isolation bag: comparison of perioperative outcomes. J Minim Invasive Gynecol 2015;22(3):433-438. [doi: 10.1016/j.jmig.2014.11.010] [Medline: 25452122]

12. Matta R, Doiron C, Leveridge MJ. The dramatic increase in social media in urology. J Urol 2014 Aug;192(2):494-498. [doi: 10.1016/j.juro.2014.02.043] [Medline: 24576656]
13. Foroughi F, Lam AK, Lim MSC, Saremi N, Ahmadvand A. Googling for cancer: an infodemiological assessment of online search interests in Australia, Canada, New Zealand, the United Kingdom, and the United States. JMIR Cancer 2016 May 04;2(1):e5 [FREE Full text] [doi: 10.2196/cancer.5212] [Medline: 28410185]
14. Stone BV, Forde JC, Levit VB, Lee RK, Te AE, Chughtai B. Trends in internet search activity, media coverage, and patient-centered health information after the FDA safety communications on surgical mesh for pelvic organ prolapse. Int Urogynecol J 2016 Nov;27(11):1761-1766. [doi: 10.1007/s00192-016-3040-6] [Medline: 27209310]
15. Crabb RM, Rafie S, Weingardt KR. Health-related internet use in older primary care patients. Gerontology 2012;58(2):164-170. [doi: 10.1159/000329340] [Medline: 21734360]

Abbreviations

FDA: US Food and Drug Administration
RSV: relative search value

Edited by G Eysenbach; submitted 22.01.18; peer-reviewed by K Eilber, M Canda; comments to author 15.02.18; revised version received 21.02.18; accepted 23.02.18; published 26.04.18.

Please cite as:
Wood LN, Jamnagerwalla J, Markowitz MA, Thum DJ, McCarty P, Medendorp AR, Raz S, Kim JH
Public Awareness of Uterine Power Morcellation Through US Food and Drug Administration Communications: Analysis of Google Trends Search Term Patterns
JMIR Public Health Surveill 2018;4(2):e47
URL: http://publichealth.jmir.org/2018/2/e47/
doi:10.2196/publichealth.9913
PMID:29699965

Original Paper

Monitoring Interest in Herpes Zoster Vaccination: Analysis of Google Search Data

Elyse J Berlinberg[1], BS; Michael S Deiner[1,2], PhD; Travis C Porco[1,2,3,4], PhD, MPH; Nisha R Acharya[1,2,3,4], MD, MS

[1]Francis I Proctor Foundation, University of California, San Francisco, San Francisco, CA, United States
[2]Department of Ophthalmology, University of California, San Francisco, San Francisco, CA, United States
[3]Department of Epidemiology & Biostatistics, University of California, San Francisco, San Francisco, CA, United States
[4]Institute for Global Health Sciences, University of California, San Francisco, San Francisco, CA, United States

Corresponding Author:
Nisha R Acharya, MD, MS
Francis I Proctor Foundation
University of California, San Francisco
513 Parnassus Ave, S309
San Francisco, CA, 94143-0412
United States
Phone: 1 415 476 8131
Fax: 1 415 502 2521
Email: nisha.acharya@ucsf.edu

Abstract

Background: A new recombinant subunit vaccine for herpes zoster (HZ or shingles) was approved by the United States Food and Drug Administration on October 20, 2017 and is expected to replace the previous live attenuated vaccine. There have been low coverage rates with the live attenuated vaccine (Zostavax), ranging from 12-32% of eligible patients receiving the HZ vaccine.

Objective: This study aimed to provide insight into trends and potential reasons for interest in HZ vaccination.

Methods: Internet search data were queried from the Google Health application programming interface from 2004-2017. Seasonality of normalized search volume was analyzed using wavelets and Fisher's g test.

Results: The search terms "shingles vaccine," "zoster vaccine," and "zostavax" all exhibited significant periodicity in the fall months (*P*<.001), with sharp increases after recommendations for vaccination by public health-related organizations. Although the terms "shingles blisters," "shingles itch," "shingles rash," "skin rash," and "shingles medicine" exhibited statistically significant periodicities with a seasonal peak in the summer (*P*<.001), the terms "shingles contagious," "shingles pain," "shingles treatment," and "shingles symptoms" did not reveal an annual trend.

Conclusions: There may be increased interest in HZ vaccination during the fall and after public health organization recommendations are broadcast. This finding points to the possibility that increased awareness of the vaccine through public health announcements could be evaluated as a potential intervention for increasing vaccine coverage.

(JMIR Public Health Surveill 2018;4(2):e10180) doi:10.2196/10180

KEYWORDS

herpes zoster; vaccination; Internet; periodicity; Google Trends; infodemiology

Introduction

One in three Americans will experience shingles (herpes zoster; HZ) in their lifetime, which may lead to extremely painful postherpetic neuralgia [1]. HZ of the eye, or herpes zoster ophthalmicus, is the second most common location of HZ and can lead to blindness [1-4]. On October 20, 2017, the United States Food and Drug Administration (FDA) approved Shingrix (GlaxoSmithKline), a new recombinant subunit vaccine for HZ [5]. The recombinant subunit vaccine has been shown to have over 95% efficacy in clinical trials and is now Centers for Disease Control and Prevention (CDC)-preferred over the live attenuated vaccine Zostavax (Merck) released in 2006 [6,7]. However, many lessons can be learned from this first experience with HZ vaccination. Despite showing over 51% and 70% efficacy in adults over age 50 and 60, respectively, the live attenuated vaccine had low coverage rates: studies estimate between 12% and 32% of vaccine-eligible adults have received the HZ vaccine [8-10]. Many possible contributing factors to the low vaccination rate have been proposed, including high cost, a lack of consensus on vaccine recommendation, waning efficacy, supply issues, and general lack of awareness of the vaccine and its benefits [11-20]. Influences on general awareness and interest in the HZ vaccine are poorly understood, and if known they could be utilized in designing more effective public health efforts to increase HZ immunizations. We sought to identify some of the factors affecting public interest in HZ and its vaccine.

Over the past several years, analysis of social media and search data has been of increasing interest in public health efforts [21,22]. This approach has been used to estimate disease incidence, identify and predict potential epidemic outbreaks, assess general understanding of a disease, and gauge public interest in a health-related topic [23-33]. In this study, we analyzed Google Search data to assess past trends in public interest for HZ vaccination in an attempt to better understand what factors may impact HZ vaccine utilization. These findings can inform possible future interventions that could help increase HZ vaccine awareness and uptake for eligible patients.

Methods

Search data were obtained from the Google Health application programming interface (API; Alphabet, Inc) using specific search terms related to HZ or HZ vaccination (“shingles vaccine,” “zoster vaccine,” “Zostavax,” “medicare vaccine,” “shingrix,” “shingles blisters,” “shingles itch,” “shingles rash,” “skin rash,” “shingles medicine,” “eye shingles,” “shingles contagious,” “shingles pain,” “shingles treatment,” and “shingles symptoms”) and negative control search terms (“roof shingle,” “roof shingles,” “roof repair,” and “roof symptoms”). Unless otherwise specified, all API queries limited results to searches made with a United States geolocation. For each term, the Google Health API was queried for all search data between January 1, 2004 and December 1, 2017. Weekly 2017 data were obtained from the public Google Trends site, which automatically normalizes searches as a relative volume, reported as a number between 0 and 1. Google API data is presented in a grouped fashion to protect privacy per Google API data user requirements.

We assessed periodicity in searches using the weekly or monthly relative intensity of the search data queried between January 1, 2011 and December 31, 2016. Significance of periodicity was tested using Fisher’s g test [34]. To compare seasonal timing between different search terms, we fit trigonometric regression models. We used the equation:

and used time series bootstrap [35] to test the null hypothesis that the difference in the phase angle arctan for the annual terms was zero. A fixed width of 4 was chosen for bootstrap resampling of residuals, with sensitivity analysis using a width of 2 and 8.

Time-frequency analysis was conducted using Morlet wavelets [36], which were estimated using the WaveletComp package in R. A *P*-value of less than .05 was considered significant. All statistical tests were conducted in R version 3.3 for MacIntosh (The R Foundation for Statistical Computing, Vienna, Austria).

Results

Seasonal Patterns of Searches Related to Herpes Zoster Vaccination

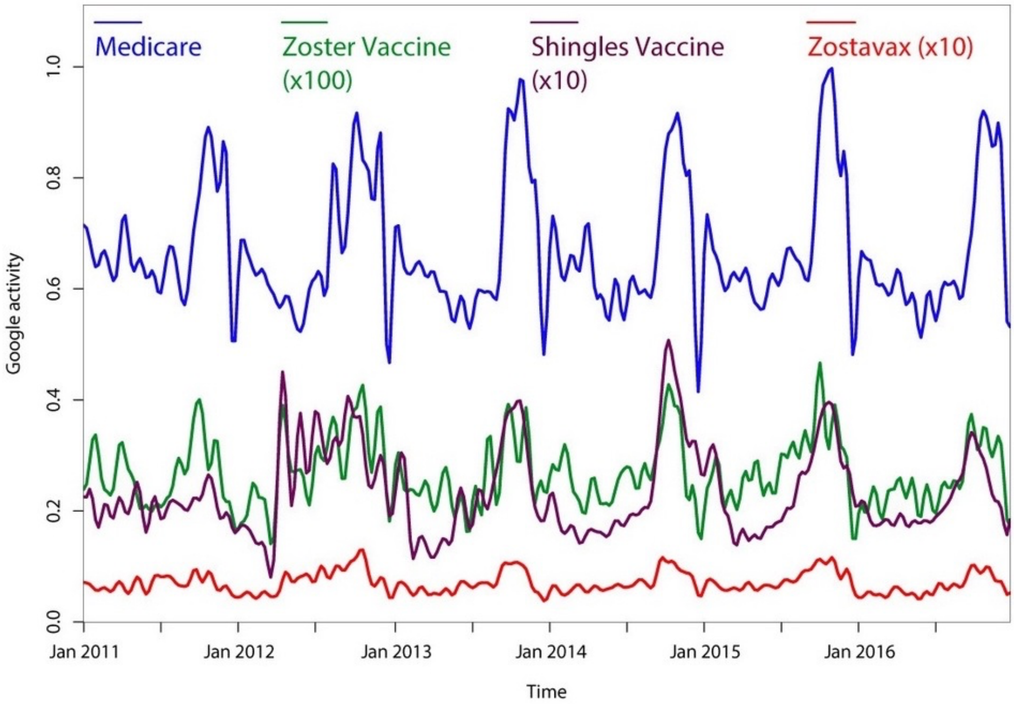
Google Health API was queried for all searches related to HZ vaccination between January 1, 2011 and December 31, 2016. The relative frequency of searches for the terms “shingles vaccine,” “zoster vaccine,” and “zostavax” are shown in Figure 1. Trend lines for these terms exhibited a significant annual periodicity with a maximum in the fall ($P<.001$ by Fisher’s g test).

Our initial hypothesis on the possible cause of this seasonal trend is that it is related to insurance utilization. Since deductibles are generally reset on January 1st in American health insurance systems, it could be that more HZ vaccine-eligible patients would seek out this elective immunization in the last few months of the year. Furthermore, the majority of patients receiving the HZ vaccine are Medicare-eligible. The open enrollment period for Medicare Advantage and the Medicare Drug Plan runs between mid-October and December [37]. Newly eligible patients may look online for ways to apply their new insurance policy immediately after they enroll. To test this hypothesis, we assessed the seasonality of searches for “medicare” (Figure 1, blue) and “medicare vaccine” (data not shown). These terms revealed a similar annual cycle as “shingles vaccine,” “zoster vaccine,” and “Zostavax,” with a significant seasonal trend peaking in the fall ($P<.001$ for both terms, Fisher’s g test).

In addition, it has been shown that covaccination for influenza and HZ is safe [38]. One could also hypothesize that the surge in interest for the HZ vaccine may be linked to the flu season. We assessed periodicity for searches for “flu vaccine” and found that it also exhibited a pronounced annual cycle ($P<.001$, Fisher’s g test) with a peak in the fall (data not shown).

We also considered that the seasonal pattern may be related to commercial advertising for the vaccine. Merck produced two widely viewed commercials featuring a popular former National Football League quarterback, initially released in September 2014 and 2015 [39,40]. These commercials may have led individuals to then search online for more information about the vaccine. However, this does not provide a complete explanation of the seasonal trend; Merck has also released other shingles-related commercials in February, April, May, June, July, and August.

Figure 1. Smoothed Google searches for herpes zoster vaccination-related terms from 2011-2017, normalized as relative search volume (0-1.0). Smoothing was conducted using Morlet wavelet denoising (lower period of 4 weeks, upper period of 128 weeks, omitting terms nonsignificant at the P=.05 level).



Significant Events in Herpes Zoster Vaccination

In our analysis for a seasonal trend in HZ vaccination, we noted that there were certain time points at which the overall volume of searches for vaccination-related topics sharply increased. We inferred that these time points may correspond with key dates in the development, approval, and extension of the HZ vaccine. The FDA approved the live attenuated vaccine (Zostavax) for US patients aged 60 or older on May 25, 2006 and extended its recommendation to patients aged 50 or older on March 25, 2011 [41,42]. Furthermore, the CDC issued its recommendation for the live attenuated vaccine in adults aged 60 or older on October 25, 2006 [43]. Google Trends data were used to estimate US public interest in the vaccine before and after these key dates (Figure 2, red). Between April 2006 and May 2006, Google searches for the term "zostavax" increased 16-fold. Google searches for "zostavax" reached a maximum in November 2006, shortly after the CDC's recommendation was broadcast. Google searches for "zostavax" decreased over time after the FDA and CDC recommendations and increased again around the date of the FDA extension in March 2011, when searches increased 1.5-fold over the subsequent 6 months and 2-fold in the following year.

The trend in searches for "zostavax" in the United States was compared to the trend in Canada, where the National Advisory Committee on Immunization (NACI) first recommended the live attenuated vaccine for patients 60 years of age or older in January 2010, and then extended this recommendation to patients 50 years of age or older in January 2014 [44,45]. In Canada (Figure 2, green), searches for "zostavax" increased 2-fold just before the initial offering of the vaccine in 2010, but their maximum did not occur until 2014, when the NACI extended its recommendation to people between 50-59 years of age. US searches for "zostavax" were also compared to searches in the United Kingdom, where National Health Service England first offered the live attenuated vaccine through its *Shingles Immunisation Programme* in September 2013 [46]. Google searches for "zostavax" in the United Kingdom (Figure 2, blue) reached a maximum in October 2013.

The FDA approved the new recombinant subunit vaccine (Shingrix) on October 20, 2017 [5]. In light of this recent event, we additionally analyzed Google searches before and after this approval. The results can be found in Figure 3. After FDA approval, Google searches for the term "shingrix" increased over 6-fold with an accompanying 2-fold increase in searches for the term "shingles vaccine" and 1.5-fold increase in searches for "zoster vaccine."

Findings for Other Herpes Zoster-Related Terms

While this study primarily focused on trends in HZ vaccination, we also used the same techniques to explore trends for other HZ-related terms that are less directly associated with vaccination, including "shingles contagious " and "shingles medicine" (Figure 4); "shingles symptoms," "shingles blisters," "shingles rash," and "eye shingles" (Figure 5); and "skin rash," "shingles itch," "shingles pain," and "shingles treatment" (data not shown). Figure 4 also illustrates trends in searches for the possible covariate "roof shingles" that also exhibits an annual periodicity but has a slightly shifted peak season compared to the terms "shingles contagious" and "shingles medicine" ($P<.001$). Although we did not find sharp increases in searches for these other HZ-related terms, we did find a periodicity for many of them. The terms "shingles blisters," "shingles itch," "shingles medicine," "skin rash," and "shingles rash" showed a statistically significant annual trend (see Table 1). The terms "shingles contagious," "eye shingles," and "shingles pain" showed no evidence of periodicity. The term “shingles symptoms” showed a highly significant periodicity with a 20-week cycle ($P<.001$). The term “shingles treatment” showed a similar 20-week periodicity that did not reach significance ($P=.11$) according to Fisher’s g test (although Morlet wavelet decomposition revealed a highly significant periodicity which began in early 2012).

Figure 2. Trends in Google searches for herpes zoster vaccination-related terms from 2004-2017 for the United States, Canada, and the United Kingdom, shown using 3-point median smooth. Trends are annotated for key dates in live attenuated vaccine recommendation.

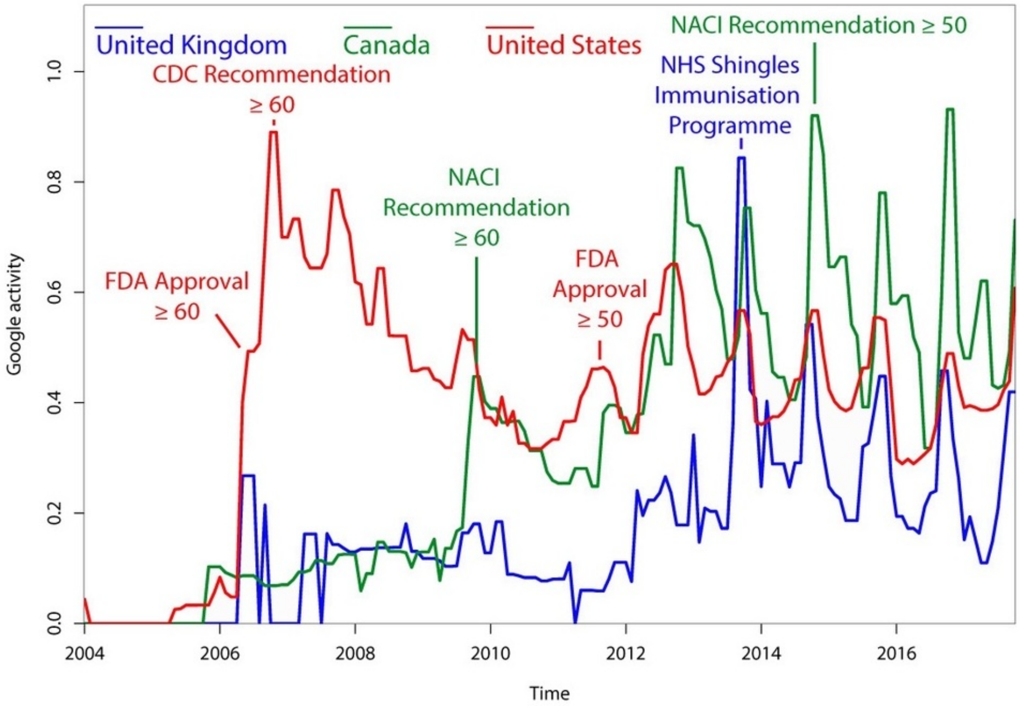


Figure 3. Google search data for herpes zoster vaccination-related terms and the term “shingrix” around its United States Food and Drug Administration approval on October 20, 2017. The 2017 weekly data used for this figure was retrieved from the public Google Trends website, not the Google Health application programming interface.

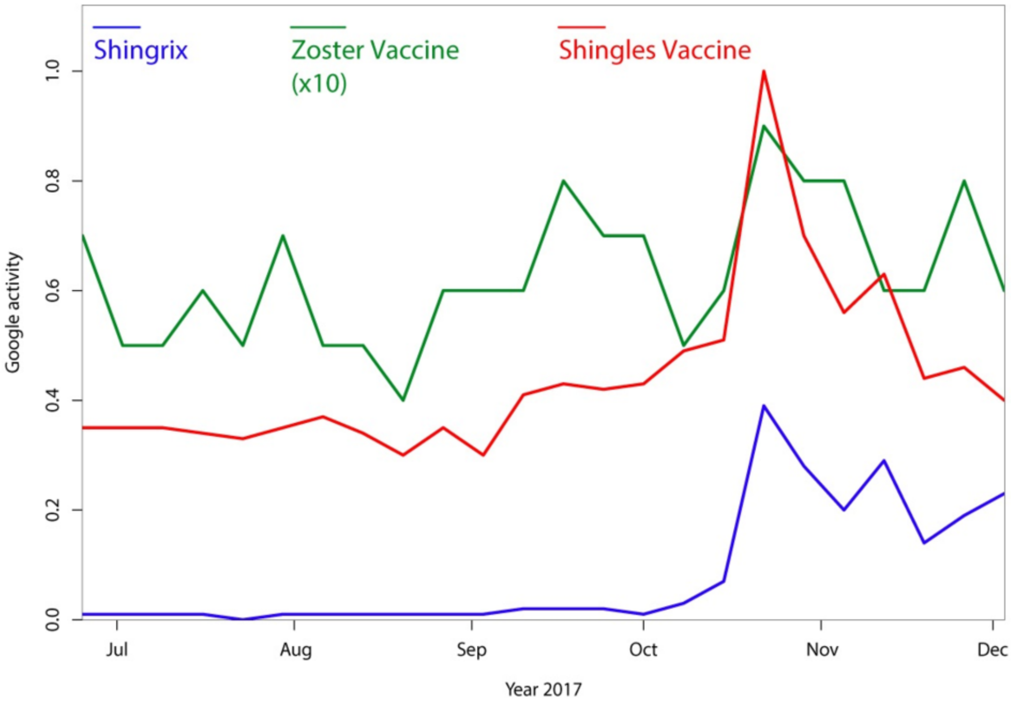


Figure 4. Smoothed Google searches for “shingles contagious,” “roof shingles,” and “shingles medicine” from 2011-2017, normalized as relative search volume (0-1.0). Searches for “roof shingles” are shown to illustrate seasonality of an unrelated search for “shingles.” Smoothing was conducted as in Figure 1.

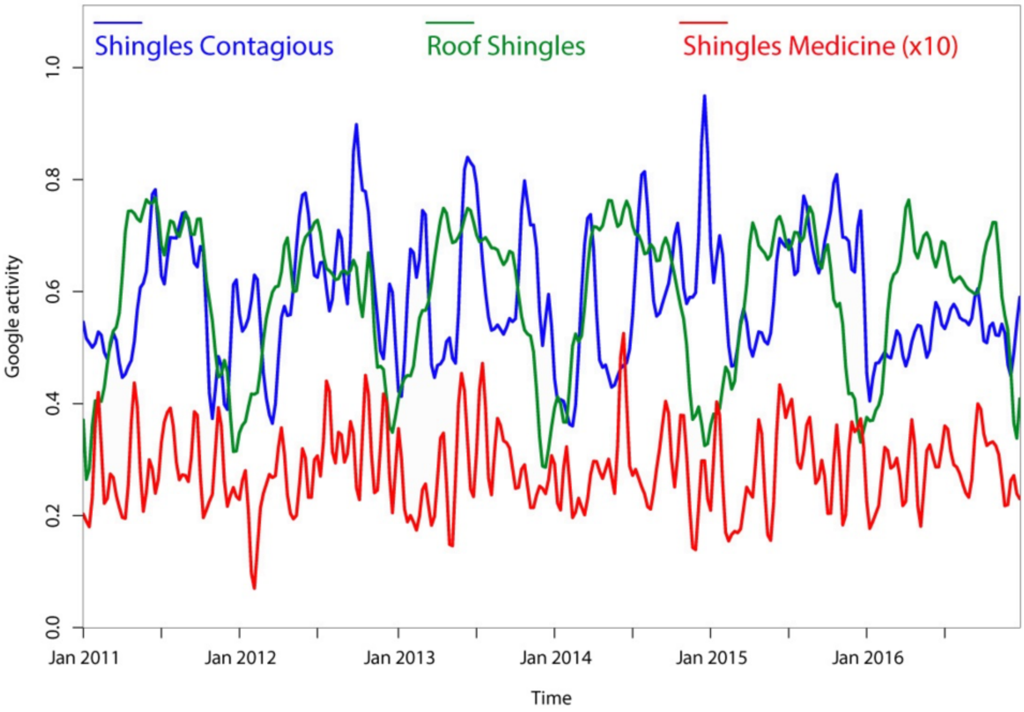


Figure 5. Smoothed Google searches for “shingles symptoms,” “shingles blisters,” “shingles rash,” and “eye shingles” from 2011-2017, normalized as relative search volume (0-1.0).

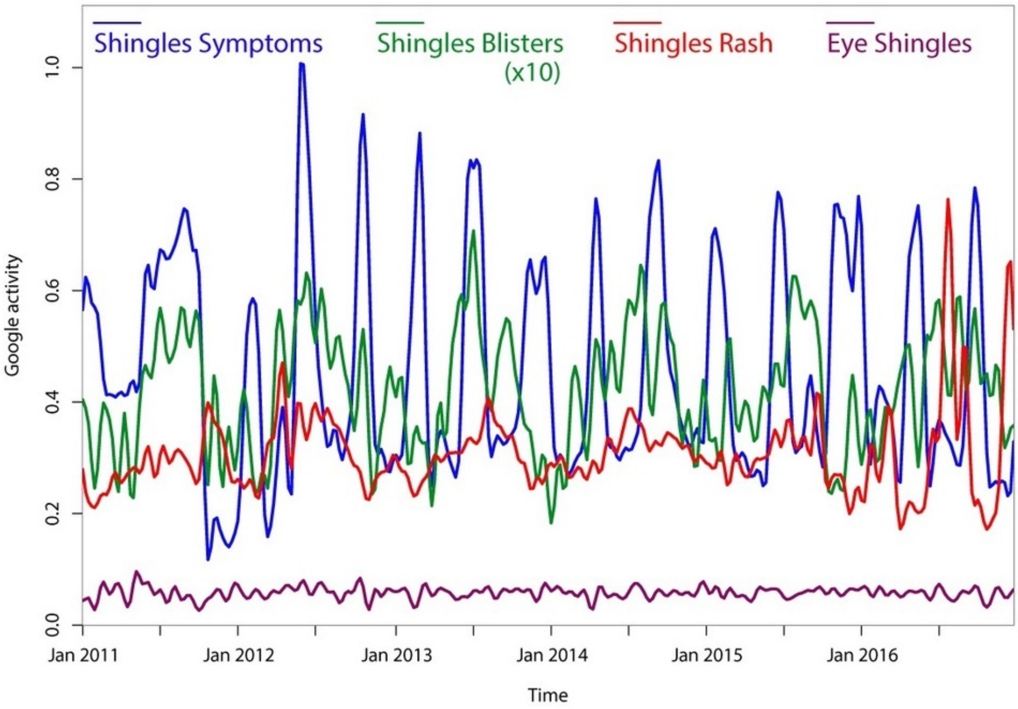


Table 1. Periodicity of searches for herpes zoster-related terms.

Search Term	Frequency (cycles per year)	Peak Season(s)	Fisher's g
Shingles Blisters	1	Summer	<.001
Shingles Itch	1	Summer	<.001
Shingles Rash	1	Summer	<.001
Skin Rash	1	Summer	<.001
Shingles Medicine	1	Summer	<.001
Eye Shingles	1	N/A[a]	.88
Shingles Contagious	<1	N/A	<.001
Shingles Pain	<1	N/A	.01
Shingles Treatment	2.6	N/A	.11
Shingles Symptoms	2.6	N/A	<.001

[a] N/A: not applicable.

Discussion

Principal Findings

Our current analysis of historic changes and seasonal patterns of public search interest over time related to the HZ vaccine identified a highly significant annual trend for searches related to vaccination, which appeared in part to be related to utilization of Medicare benefits, covaccination with the influenza vaccine, and advertising campaigns. This finding provides us with a better understanding of public interest in the HZ vaccine, which can be applied to informing future efforts to increase vaccine coverage.

Our results also suggest that public interest in a vaccine may be significantly affected by recommendations from health-related governmental departments. In the United States, United Kingdom, and Canada, a spike in search interest occurred in each country for vaccine-related terms soon after a recommendation by a health-related governmental organization was broadcast. With this in mind, it may be helpful to increase vaccine production or clinical stocks in anticipation of the increase in vaccine interest caused by these announcements. This factor is particularly important given the recent release of the new recombinant subunit vaccine. When the live attenuated vaccine was first released, there was some delay in vaccination efforts due to an insufficient supply of the vaccine [13]. The recent release of the recombinant subunit vaccine should not fall into the same trap, and vaccine availability could be increased accordingly to anticipate the likely increase in interest after any FDA or CDC announcements.

Finally, we report on a possible seasonality for interest in HZ as measured by an increase in Google search intensity in the summer. While trends in Google searches could be representative of trends in HZ incidence, there have been mixed reports about whether or not HZ is seasonal. Many studies have been published on the topic, ranging from reports of a single practice to regional health registries [47-58]. Six of these studies found that HZ exhibits a statistically significant seasonality, while two found that HZ incidence varies by season, but this trend was either not significant or its significance was not tested. All of the studies that reported a seasonal trend found an increase of HZ cases in the summer, which could explain the increase in interest for HZ vaccination in the summer. However, Google searches are only representative of interest in a topic and do not necessarily mean that a person has the disease they have searched for. Since the seasonality of HZ has been inconsistently characterized, we believe that this trend should be evaluated further utilizing a large health-specific dataset.

Limitations

The use of social media and Internet search data to assess health matters has become more useful in recent years, but its methods come with limitations including the following for our current study. Google API data are reported as a normalized relative volume and can over-represent fluctuations in searches for terms with low post volumes. Second, Google Searches also estimate interest in a given topic, but not the follow through of getting vaccinated. Likewise, the periodicity of searches for HZ vaccination may not correspond to trends in actual utilization of the vaccine. As mentioned previously, our findings could benefit from validation using large administrative clinical datasets.

Conclusions

Google searches about HZ vaccination exhibit an annual periodicity in the fall and significant peaks after recommendations from large public health organizations are broadcast. Given previous issues with HZ vaccine availability and coverage, these trends may help predict when to increase supply of the new recombinant subunit vaccine to meet changes in demand, and when to increase awareness to improve coverage.

Acknowledgments

This work was supported in part by grants 1R01EY024608 and 1R01EY028739 from the National Institutes of Health National Eye Institute (NIH-NEI), grant EY002162 (Core Grant for Vision Research) from the NIH-NEI, and an Unrestricted Grant from Research to Prevent Blindness.

Conflicts of Interest

None declared

References

1. Dworkin RH, Johnson RW, Breuer J, Gnann JW, Levin MJ, Backonja M, et al. Recommendations for the management of herpes zoster. Clin Infect Dis 2007 Jan 01;44 Suppl 1:S1-26. [doi: 10.1086/510206] [Medline: 17143845]
2. Liesegang TJ. Herpes zoster ophthalmicus natural history, risk factors, clinical presentation, and morbidity. Ophthalmology 2008 Feb;115(2 Suppl):S3-12. [doi: 10.1016/j.ophtha.2007.10.009] [Medline: 18243930]
3. Borkar DS, Tham VM, Esterberg E, Ray KJ, Vinoya AC, Parker JV, et al. Incidence of herpes zoster ophthalmicus: results from the Pacific Ocular Inflammation Study. Ophthalmology 2013 Mar;120(3):451-456 [FREE Full text] [doi: 10.1016/j.ophtha.2012.09.007] [Medline: 23207173]
4. Ragozzino MW, Melton LJ, Kurland LT, Chu CP, Perry HO. Population-based study of herpes zoster and its sequelae. Medicine (Baltimore) 1982 Sep;61(5):310-316. [Medline: 6981045]
5. Gruber M. United States Food and Drug Administration. 2017 Oct 20. BLA approval - zoster vaccine recombinant URL: https://www.fda.gov/downloads/BiologicsBloodVaccines/Vaccines/ApprovedProducts/UCM581750.pdf [accessed 2017-12-16] [WebCite Cache ID 6xxvVGFcq]
6. Lal H, Cunningham AL, Godeaux O, Chlibek R, Diez-Domingo J, Hwang S, ZOE-50 Study Group. Efficacy of an adjuvanted herpes zoster subunit vaccine in older adults. N Engl J Med 2015 May 28;372(22):2087-2096. [doi: 10.1056/NEJMoa1501184] [Medline: 25916341]
7. United States Centers for Disease Control and Prevention. 2018. Shingles (herpes zoster) URL: https://www.cdc.gov/shingles/vaccination.html [accessed 2018-02-02] [WebCite Cache ID 6yYPw24AY]
8. Williams WW, Lu P, O'Halloran A, Kim DK, Grohskopf LA, Pilishvili T, Centers for Disease Control and Prevention. Surveillance of vaccination coverage among adult populations - United States, 2014. MMWR Surveill Summ 2016 Feb 5;65(1):1-36. [doi: 10.15585/mmwr.ss6501a1] [Medline: 26844596]
9. Lu P, O'Halloran A, Williams WW, Harpaz R. National and state-specific shingles vaccination among adults aged ≥60 Years. Am J Prev Med 2017 Mar;52(3):362-372 [FREE Full text] [doi: 10.1016/j.amepre.2016.08.031] [Medline: 27720342]
10. Hernandez PO, Javed S, Mendoza N, Lapolla W, Hicks LD, Tyring SK. Family history and herpes zoster risk in the era of shingles vaccination. J Clin Virol 2011 Dec;52(4):344-348. [doi: 10.1016/j.jcv.2011.08.014] [Medline: 21924676]
11. Harpaz R, Ortega-Sanchez IR, Seward JF, Advisory Committee on Immunization Practices (ACIP) Centers for Disease Control and Prevention (CDC). Prevention of herpes zoster: recommendations of the Advisory Committee on Immunization Practices (ACIP). MMWR Recomm Rep 2008 Jun 06;57(RR-5):1-30 [FREE Full text] [Medline: 18528318]
12. Hales CM, Harpaz R, Ortega-Sanchez I, Bialek SR, Centers for Disease Control and Prevention (CDC). Update on recommendations for use of herpes zoster vaccine. MMWR Morb Mortal Wkly Rep 2014 Aug 22;63(33):729-731 [FREE Full text] [Medline: 25144544]
13. United States Centers for Disease Control and Prevention. 2007 May 11. Morbidity and mortality weekly report. Notice to readers: supply of vaccines containing varicella-zoster virus URL: https://www.cdc.gov/mmwr/preview/mmwrhtml/mm5618a6.htm [accessed 2017-12-21] [WebCite Cache ID 6xNbmcUpo]
14. Cohen E. Prevention of herpes zoster: we need to do better. JAMA Ophthalmol 2013 Mar;131(3):396-398. [doi: 10.1001/jamaophthalmol.2013.588] [Medline: 23494045]
15. Jung JJ, Elkin ZP, Li X, Goldberg JD, Edell AR, Cohen MN, et al. Increasing use of the vaccine against zoster through recommendation and administration by ophthalmologists at a city hospital. Am J Ophthalmol 2013 May;155(5):787-795. [doi: 10.1016/j.ajo.2012.11.022] [Medline: 23394910]
16. Hurley LP, Bridges CB, Harpaz R, Allison MA, O'Leary ST, Crane LA, et al. Physician attitudes toward adult vaccines and other preventive practices, United States, 2012. Public Health Rep 2016;131(2):320-330 [FREE Full text] [doi: 10.1177/003335491613100216] [Medline: 26957667]
17. Otsuka SH, Tayal NH, Porter K, Embi PJ, Beatty SJ. Improving herpes zoster vaccination rates through use of a clinical pharmacist and a personal health record. Am J Med 2013 Sep;126(9):832.e1-832.e6 [FREE Full text] [doi: 10.1016/j.amjmed.2013.02.018] [Medline: 23830534]
18. Bryan AR, Liu Y, Kuehl PG. Advocating zoster vaccination in a community pharmacy through use of personal selling. J Am Pharm Assoc (2003) 2013;53(1):70-77. [Medline: 23636159]
19. Opstelten W, van Essen GA, Hak E. Determinants of non-compliance with herpes zoster vaccination in the community-dwelling elderly. Vaccine 2009 Jan 07;27(2):192-196. [doi: 10.1016/j.vaccine.2008.10.047] [Medline: 18996427]

20. Teeter BS, Garza KB, Stevenson TL, Williamson MA, Zeek ML, Westrick SC. Factors associated with herpes zoster vaccination status and acceptance of vaccine recommendation in community pharmacies. Vaccine 2014 Sep 29;32(43):5749-5754. [doi: 10.1016/j.vaccine.2014.08.040] [Medline: 25171848]
21. Eysenbach G. Infodemiology and infoveillance: tracking online health information and cyberbehavior for public health. Am J Prev Med 2011 May;40(5 Suppl 2):S154-S158. [doi: 10.1016/j.amepre.2011.02.006] [Medline: 21521589]
22. Ginsberg J, Mohebbi MH, Patel RS, Brammer L, Smolinski MS, Brilliant L. Detecting influenza epidemics using search engine query data. Nature 2009 Feb 19;457(7232):1012-1014. [doi: 10.1038/nature07634] [Medline: 19020500]
23. Deiner MS, Lietman TM, McLeod SD, Chodosh J, Porco TC. Surveillance tools emerging from search engines and social media data for determining eye disease patterns. JAMA Ophthalmol 2016 Sep 01;134(9):1024-1030 [FREE Full text] [doi: 10.1001/jamaophthalmol.2016.2267] [Medline: 27416554]
24. Kang M, Zhong H, He J, Rutherford S, Yang F. Using Google Trends for influenza surveillance in South China. PLoS One 2013;8(1):e55205 [FREE Full text] [doi: 10.1371/journal.pone.0055205] [Medline: 23372837]
25. Hiller KM, Stoneking L, Min A, Rhodes SM. Syndromic surveillance for influenza in the emergency department-a systematic review. PLoS One 2013;8(9):e73832 [FREE Full text] [doi: 10.1371/journal.pone.0073832] [Medline: 24058494]
26. Wang H, Chen D, Yu H, Chen Y. Forecasting the incidence of dementia and dementia-related outpatient visits with google trends: evidence from Taiwan. J Med Internet Res 2015 Nov 19;17(11):e264 [FREE Full text] [doi: 10.2196/jmir.4516] [Medline: 26586281]
27. Dugas AF, Jalalpour M, Gel Y, Levin S, Torcaso F, Igusa T, et al. Influenza forecasting with Google Flu Trends. PLoS One 2013;8(2):e56176 [FREE Full text] [doi: 10.1371/journal.pone.0056176] [Medline: 23457520]
28. Malik MT, Gumel A, Thompson LH, Strome T, Mahmud SM. “Google flu trends” and emergency department triage data predicted the 2009 pandemic H1N1 waves in Manitoba. Can J Public Health 2011;102(4):294-297. [Medline: 21913587]
29. Schootman M, Toor A, Cavazos-Rehg P, Jeffe DB, McQueen A, Eberth J, et al. The utility of Google Trends data to examine interest in cancer screening. BMJ Open 2015 Jun 08;5(6):e006678 [FREE Full text] [doi: 10.1136/bmjopen-2014-006678] [Medline: 26056120]
30. Moccia M, Palladino R, Falco A, Saccà F, Lanzillo R, Brescia MV. Google Trends: new evidence for seasonality of multiple sclerosis. J Neurol Neurosurg Psychiatry 2016 Dec;87(9):1028-1029. [doi: 10.1136/jnnp-2016-313260] [Medline: 27083532]
31. Bragazzi NL, Bacigaluppi S, Robba C, Nardone R, Trinka E, Brigo F. Infodemiology of status epilepticus: a systematic validation of the Google Trends-based search queries. Epilepsy Behav 2016 Feb;55:120-123. [doi: 10.1016/j.yebeh.2015.12.017] [Medline: 26773681]
32. Pollett S, Wood N, Boscardin WJ, Bengtsson H, Schwarcz S, Harriman K, et al. Validating the use of Google Trends to enhance pertussis surveillance in California. PLoS Curr 2015 Oct 19;7 [FREE Full text] [doi: 10.1371/currents.outbreaks.7119696b3e7523faa4543faac87c56c2] [Medline: 26543674]
33. Foroughi F, Lam AK, Lim MSC, Saremi N, Ahmadvand A. “Googling” for cancer: an infodemiological assessment of online search interests in Australia, Canada, New Zealand, the United Kingdom, and the United States. JMIR Cancer 2016 May 04;2(1):e5 [FREE Full text] [doi: 10.2196/cancer.5212] [Medline: 28410185]
34. Fisher RA. Tests of significance in harmonic analysis. Proc Royal Soc A 1929 Aug 01;125(796):54-59. [doi: 10.1098/rspa.1929.0151]
35. Kreiss J, Lahiri S. Bootstrap methods for time series. In: Subba Rao T, Subba Rao S, Rao CR, editors. Time Series Analysis: Methods and Applications, Volume 30. North Holland: Elsevier; 2012:3-26.
36. Pereyra M, Ward L. Harmonic Analysis: From Fourier to Wavelets. Park City, UT: American Mathematical Society; 2012.
37. United States Centers for Medicare & Medicaid Services. 2018. Joining a health or drug plan URL: https://www.medicare.gov/sign-up-change-plans/ [accessed 2018-04-09] [WebCite Cache ID 6yYQTuWAi]
38. Kerzner B, Murray AV, Cheng E, Ifle R, Harvey PR, Tomlinson M, et al. Safety and immunogenicity profile of the concomitant administration of ZOSTAVAX and inactivated influenza vaccine in adults aged 50 and older. J Am Geriatr Soc 2007 Oct;55(10):1499-1507. [doi: 10.1111/j.1532-5415.2007.01397.x] [Medline: 17908055]
39. iSpot.tv. 2014. Merck TV commercial, 'Football Legend' featuring Terry Bradshaw URL: https://www.ispot.tv/ad/7CpH/merck-football-legend-featuring-terry-bradshaw [accessed 2018-03-16] [WebCite Cache ID 6xxulxb9l]
40. iSpot.tv. 2015. Merck TV commercial, 'Surprise Door Knock' featuring Terry Bradshaw URL: https://www.ispot.tv/ad/7UEu/merck-surprise-door-knock-featuring-terry-bradshaw [accessed 2018-03-16] [WebCite Cache ID 6xxuuHAYn]
41. Baylor N. United States Food and Drug Administration. 2006 May 25. Approval Letter - Zostavax URL: http://wayback.archive-it.org/7993/20170723093336/https://www.fda.gov/BiologicsBloodVaccines/Vaccines/ApprovedProducts/ucm132873.htm [accessed 2017-12-03] [WebCite Cache ID 6xxv4LuJy]
42. Agger PE, Sun W, United States Food and Drug Administration. Summary Basis for Regulatory Action. 2011. URL: https://wayback.archive-it.org/7993/20170404184105/https://www.fda.gov/downloads/BiologicsBloodVaccines/Vaccines/ApprovedProducts/UCM249230.pdf [accessed 2018-03-16] [WebCite Cache ID 6xxvBM1cQ]
43. United States Centers for Disease Control and Prevention. 2006. Press release: CDC's advisory committee recommends “shingles” vaccination URL: https://www.cdc.gov/media/pressrel/r061026.htm [accessed 2018-02-20] [WebCite Cache ID 6xNc2woaD]

44. National Advisory Committee on Immunization. An Advisory Committee Statement (ACS). National Advisory Committee on Immunization (NACI). Statement on recommended use of pneumococcal conjugate vaccine. Can Commun Dis Rep 2002 Jan 15;28(ACS-2):1-32 [FREE Full text] [Medline: 12728645]
45. National Advisory Committee on Immunization. Public Health Agency of Canada. 2014 Jan. An Advisory Committee Statement (ACS) National Advisory Committee on Immunization (NACI): Update on the Use of the Herpes Zoster Vaccine URL: https://www.canada.ca/content/dam/phac-aspc/documents/services/publications/healthy-living/update-use-herpes-zoster-vaccine/hzv-vcz-eng.pdf [accessed 2017-12-10] [WebCite Cache ID 6yYOBsAqc]
46. Immunisation Implementation & Planning Public Health England. Shingles Immunisation Programm. 2013 Apr. Public health functions to be exercised by NHS England: Service Specification URL: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/192971/S7A_VARIATION_2013-14_FINAL_130417.pdf [accessed 2017-12-03] [WebCite Cache ID 6xxu39g02]
47. Glynn C, Crockford G, Gavaghan D, Cardno P, Price D, Miller J. Epidemiology of shingles. J R Soc Med 1990 Oct;83(10):617-619 [FREE Full text] [Medline: 1962821]
48. Gallerani M, Manfredini R. Correspondence: seasonal variation in herpes zoster infection. Br J Dermatol 2000;142:588. [doi: 10.1046/j.1365-2133.2000.03400.x]
49. Termorshuizen F, Hogewoning AA, Bouwes BJN, Goettsch WG, de Fijter JW, van Loveren H. Skin infections in renal transplant recipients and the relation with solar ultraviolet radiation. Clin Transplant 2003 Dec;17(6):522-527. [Medline: 14756268]
50. Wu P, Wu HI, Chou T, Sung F. Varicella vaccination alters the chronological trends of herpes zoster and varicella. PLoS One 2013;8(10):e77709 [FREE Full text] [doi: 10.1371/journal.pone.0077709] [Medline: 24204928]
51. Zak-Prelich M, Borkowski JL, Alexander F, Norval M. The role of solar ultraviolet irradiation in zoster. Epidemiol Infect 2002 Dec;129(3):593-597. [Medline: 12558343]
52. Jardine A, Conaty SJ, Vally H. Herpes zoster in Australia: evidence of increase in incidence in adults attributable to varicella immunization? Epidemiol Infect 2011 May;139(5):658-665. [doi: 10.1017/S0950268810001949] [Medline: 20727248]
53. Hope-Simpson SE. The nature of herpes zoster: a long-term study and a new hypothesis. Proc R Soc Med 1965 Jan;58:9-20 [FREE Full text] [Medline: 14267505]
54. Toyama N, Shiraki K, Society of the Miyazaki Prefecture Dermatologists. Epidemiology of herpes zoster and its relationship to varicella in Japan: A 10-year survey of 48,388 herpes zoster cases in Miyazaki prefecture. J Med Virol 2009 Dec;81(12):2053-2058. [doi: 10.1002/jmv.21599] [Medline: 19856466]
55. Kim YJ, Lee CN, Lim C, Jeon WS, Park YM. Population-based study of the epidemiology of herpes zoster in Korea. J Korean Med Sci 2014 Dec;29(12):1706-1710 [FREE Full text] [doi: 10.3346/jkms.2014.29.12.1706] [Medline: 25469074]
56. Cooper M. The epidemiology of herpes zoster. Eye (Lond) 1987;1 (Pt 3):413-421. [doi: 10.1038/eye.1987.63] [Medline: 3653445]
57. Korostil IA, Regan DG. Varicella-zoster virus in Perth, Western Australia: seasonality and reactivation. PLoS One 2016;11(3):e0151319 [FREE Full text] [doi: 10.1371/journal.pone.0151319] [Medline: 26963841]
58. Brisson M, Edmunds WJ, Law B, Gay NJ, Walld R, Brownell M, et al. Epidemiology of varicella zoster virus infection in Canada and the United Kingdom. Epidemiol Infect 2001 Oct;127(2):305-314. [Medline: 11693508]

Abbreviations

API: application programming interface
FDA: United States Food and Drug Administration
HZ: herpes zoster
NACI: Canadian National Advisory Committee on Immunization
NIH-NEI: National Institutes of Health National Eye Institute

---

Edited by M Focsa; submitted 20.02.18; peer-reviewed by N Bragazzi, G Rutherford; comments to author 16.03.18; revised version received 19.03.18; accepted 20.03.18; published 02.05.18.

Please cite as:
Berlinberg EJ, Deiner MS, Porco TC, Acharya NR
Monitoring Interest in Herpes Zoster Vaccination: Analysis of Google Search Data
JMIR Public Health Surveill 2018;4(2):e10180
URL: http://publichealth.jmir.org/2018/2/e10180/
doi:10.2196/10180
PMID:29720364

Original Paper

New Tobacco and Tobacco-Related Products: Early Detection of Product Development, Marketing Strategies, and Consumer Interest

Yvonne CM Staal[1], PhD; Suzanne van de Nobelen[1,2], PhD; Anne Havermans[1], PhD; Reinskje Talhout[1], PhD

[1]RIVM, Centre for Health Protection, Bilthoven, Netherlands
[2]Johnson & Johnson Janssen Vaccines, Leiden, Netherlands

Corresponding Author:
Reinskje Talhout, PhD
RIVM
Centre for Health Protection
Antonie van Leeuwenhoeklaan 9
Bilthoven,
Netherlands
Phone: 31 30 274 4505
Fax: 31 30 274 4505
Email: Reinskje.Talhout@rivm.nl

Abstract

Background: A wide variety of new tobacco and tobacco-related products have emerged on the market in recent years.

Objective: To understand their potential implications for public health and to guide tobacco control efforts, we have used an infoveillance approach to identify new tobacco and tobacco-related products.

Methods: Our search for tobacco(-related) products consists of several tailored search profiles using combinations of keywords such as “e-cigarette” and “new” to extract information from almost 9000 preselected sources such as websites of online shops, tobacco manufacturers, and news sites.

Results: Developments in e-cigarette design characteristics show a trend toward customization by possibilities to adjust temperature and airflow, and by the large variety of flavors of e-liquids. Additionally, more e-cigarettes are equipped with personalized accessories, such as mobile phones, applications, and Bluetooth. Waterpipe products follow the trend toward electronic vaping. Various heat-not-burn products were reintroduced to the market.

Conclusions: Our search for tobacco(-related) products was specific and timely, though advances in product development require ongoing optimization of the search strategy. Our results show a trend toward products resembling tobacco cigarettes vaporizers that can be adapted to the consumers’ needs. Our search for tobacco(-related) products could aid in the assessment of the likelihood of new products to gain market share, as a possible health risk or as an indicator for the need on independent and reliable information of the product to the general public.

(JMIR Public Health Surveill 2018;4(2):e55) doi:10.2196/publichealth.7359

KEYWORDS

noncigarette tobacco products; electronic nicotine delivery systems; public opinion; retrospective studies

Introduction

Background

A wide variety of new tobacco- and smoking-related products have emerged on the market in recent years. Moreover, tobacco companies will keep developing new products to keep meeting the changing needs of their consumers and fulfill changing regulatory requirements. These new tobacco-related products can quickly gain popularity [1], even before there is sufficient scientific evidence to determine their effects on the user and bystander. For instance, recently products marketed as “harm reduction,” “reduced risk,” or “next generation” products [2-4] were introduced making claims of being up to 90% less harmful than conventional cigarettes. Even though these products seem very attractive to consumers, independent scientific research to support these claims is lacking. In addition, these products may be attractive for smokers, but may also be used by nonsmokers.

Besides, regardless of their own harmfulness, popular new products could also serve as a gateway to the use of tobacco or tobacco-related products.

Given their potential harmful health effects, and the possibility that the product serves as a gateway to the use of more harmful products, it is important for national authorities and scientists to closely monitor product development of new tobacco and tobacco-related products. Data on product development, marketing strategies, and consumer interest in new products could help to understand potential implications for public health and guide tobacco control efforts [5]. The World Health Organization recognizes the importance of monitoring the availability and regulation of new tobacco and tobacco-related products entering national and global markets [6].

An important determinant of the popularity of the product is product marketing. For instance, previous research has shown that product marketing plays an important role in the perception of the product for potential users of e-cigarettes [7,8]. Factors that increased the likelihood of potential users to try an e-cigarette were as follows: no health warnings, the presence of flavors, and low or no nicotine content. Similar findings have been reported for waterpipe products [9]. Smokers and recent quitters appear receptive to e-cigarettes when exposed to advertisements of the products [10]. Recently, a whole new possibility for marketing has become available through social media. Retailers can easily advertise and promote their products to a large audience through websites such as twitter and Instagram [11,12]. These marketing aspects of the products themselves and in advertisements can lead to fast increases in the popularity of products and should therefore be considered when evaluating the market developments of new products.

Objective

Retrospective mapping of media trends can provide valuable insights in the types of tobacco products that were popular in the past and can help predict which products will become popular in the future. This paper describes retrospective trends in product development of e-cigarettes, based on data collected over the period of January 2014 until May 2016. During that period, we have collected information from scientific literature and publicly available sources such as news websites and patent databases, to monitor developments in design characteristics of e-cigarettes and related products such as heat-not-burn products. To this purpose, we used an automated search for emerging tobacco(-related) products, and in parallel, we monitored public interest. We have used an approach that has some characteristics of infoveillance, that is, a type of syndromic surveillance that utilizes Web-based contents, and similar aspects as applied by others in this field, that is, using Google searches, social media content, or information from websites [13,14]. Infoveilance has been used in tobacco research to identify trends or to obtain information on perception or user groups [11,15-17]. We have used this approach to identify products that contained new or unconventional design characteristics, such as technical features and decorative elements, to gain insight in the way these products were marketed.

Methods

Product Development

To search websites to monitor product development and marketing strategies, we defined a tailored search profile consisting of specific keywords (eg, “e-cigarette” and “hookah”), which were combined with keywords indicating a development (eg, “innovat*” and “trend*”) in both Dutch and English (Multimedia Appendix 1). We selected these languages as we aimed to identify products that were new or soon to be on the Dutch market. For these products, we expect marketing to be in Dutch or English. This search profile was applied to almost 9000 internet sources, which were searched daily (Multimedia Appendix 2). These sources included national and international news websites, websites of online shops, and websites of tobacco manufacturers and wholesalers. Our search resulted in a list of internet links to websites containing information that met our selected keywords. The information on these websites was reviewed for relevancy to our aims. Irrelevant information, eg, not related to tobacco in any way, was removed. These “false” hits were used to optimize the search algorithm by excluding or adding keywords. The initial focus of our search strategy was the e-cigarette. Nevertheless, we also found messages on product development of related products such as heat-not-burn products. The automated search was subsequently extended to follow developments in different classes of products simultaneously.

When a new type of product was identified in our search, a manual internet search and scientific literature search were performed to obtain more information on specific aspects of this new product. In addition, we used product reviews, consumer specific forums, and other publicly available data such as information provided by producers. Next, a database was created containing all new products identified by our search. Products were included in the database when they were considered to have new or unconventional characteristics. This could be products that were based on a new technology, products with new applications (such as the combination with a mobile phone), products that had a new or alternative use (such as new user groups, specific target groups, unintended use), or products showing an increase in popularity (by the increase in messages in our search).

All selected products were stored in the database along with a link to the website where the particular product was found and the date on which the message was found by our search protocol. The database also included (if available) a description of the product (product type, composition, physical parameters, design features, package) and specific product features such as whether or not it contained nicotine, whether tobacco (if present) was heated or combusted, and the type of organic material to be inhaled (such as tobacco or herbs). In addition, the product’s costs relative to that of other tobacco products and its date of market release were stored. Other product marketing information was included if available, such as data on target groups (eg, young adults), reasons for use, prevalence and patterns of use, including possibilities of combined use with other products, awareness and perception of the public toward the product, and attitudes toward tobacco control policies.

It should be noted that the aim of our search was to detect trends in product development of tobacco-related products. We did not aim to detect every new model of the e-cigarette or to receive all information on a product, but instead focused on new technologies, new appearances or applications, and new properties for the user to be able to follow trends.

Public Interest

In addition, to obtain information on public interest in product groups or specific products over the course of time or in specific countries, global product information searches performed by Google were analyzed using Google trends. The numbers of searches performed for a specific term are reported in proportion to the total number of searches done on Google in that specific time period. Thus, a declining trend line indicates that the relative popularity of a search term decreases, not necessarily that the total number of searches for that term declines. It has been shown that the information obtained from Google trends matched with actual trends in popularity of e-cigarette types [18].

E-Cigarettes

Trends in Devices and Liquids

The increasing numbers of accessories on the market as well as new vaping products combined with unrelated functions attest to a clear trend toward customization of e-cigarettes. It seems that experienced users like to adopt the e-cigarette to their (inhalation) needs, leading to e-cigarettes with adjusted airflow inlet using atomizer heads with different sized air holes [19]. This is applied in the most recently introduced models, which are activated by a pressure difference when the user inhales from the e-cigarette, avoiding pressing a button to heat the device [20]. Other interesting new e-cigarette-like devices provide a combined function with other electronic products such as a Bluetooth e-cigarette, which combines vaping with listening to music or calling friends [21] and another device can be used both as e-cigarette and mobile phone [22].

Moreover, smartphone applications were introduced that track the number of e-cigarette puffs taken, calculate cost savings and increased life expectancy, and have features such as auto-shut down and password protection safety [23]. In line with this, Phillip Morris has filed a patent for an e-cigarette that is Wi-Fi connected, and thus would be able to connect to other devices [24]. This device could potentially synchronize to a smartphone application that is intended to help people quit smoking, and carefully track their progress. A similar product is the Vaporcade Jupiter, a “cellular vaporizer,” combining a smartphone with an e-cigarette [25]. This allows the user to monitor the e-cigarette use, the e-liquid remaining, and the flavor used.

Next to the expanding technical possibilities of e-cigarettes, the variation in e-liquids was growing. At the time of our search, a wide variety of e-liquids in different flavors was available on the market, such as vanilla, cocoa, cherry, tobacco, and coffee [26]. Other available e-liquids strikingly contained vitamins or cannabis flavors. In addition, specific e-cigarettes (mods) are available that allow for not only liquids but also herbs, oils, or fruits to be vaped. Moreover, dual-function devices handle both concentrates and e-liquids using multiple cartridges. We also found that manufacturers attempted to reduce the formation of formaldehyde and metal substances of vapor by producing an e-liquid in which propylene glycol is replaced by vegetable glycerin [27,28].

Waterpipe products also followed the trend of electronic vaping instead of smoking, with products appearing like the e-hookah and “waterpipe flavored” e-liquids [29-31]. Electronic waterpipe products such as e-hookah or e-shisha are highly similar to electronic cigarettes. For example, the Vitacig is a pen-shaped product that produces vapor containing flavorings and vitamins [32]. Most e-hookahs have the size of an e-cigarette, which allows the user to take the product with them easily. But there are also larger e-hookahs available, which are rechargeable, contain multiple flavor cartridges, and have a large capacity battery which can last up to 1500 puffs. Cartridges available for this e-hookah contain up to 12 ml of e-liquid [30].

Marketing, Packaging, and Labeling

To gain market share, products are advertised in multiple ways and on various platforms. On social media, products are advertised and sometimes offered for free such as batteries for e-cigarettes. Alternatively, free shipping or discounts up to 10% are offered for various products. This offering of products with reduced pricing involves dozens of messages monthly. Manufacturer or importer websites sometimes offer the possibility to become an ambassador of their product line [33]. Such an ambassador is expected to set up a community to share experiences and being active as a blogger or on social media. Some marketing activities were specifically aimed at youth by focusing on youth culture. For instance, pop-up bars set at various events and locations are part of the campaign of the e-cigarette “Juul.” The “bar” in this case is a vapor lounge with brightly colored billboard display. In addition, images or cartoons of a young woman were shown in advertisements. Other advertisements were aimed at smokers who would switch from cigarettes to e-cigarettes, describing the technology of some products as “developed by smokers,” “same feeling as a real cigarette,” or “smoking the healthy way.”

General themes that were seen in marketing of e-cigarettes are health, lifestyle, and personalization (Table 1). Wordings implicating the product is a healthier alternative are often used in marketing, such as “natural,” “food- or pharma grade,” “homeopathic,” and “made in Switzerland.” These terms refer to the plastics or e-liquid used or more specifically to the nicotine in the product. However, in most countries, worldwide restrictions on tobacco advertising are active. Before May 20, 2016, European regulations advised “to adopt a restrictive approach to advertising electronic cigarettes and refill containers” [34]. Following the Tobacco Product Directive, which became effective on May 20, 2016, advertisements or promotions for tobacco products, including e-cigarettes, are no longer permitted.

Table 1. Product appeal and marketing terms used for online e-cigarette advertisement.

Product name	Marketing terms	Product appearance
E-Njoint	Natural, harmless, safe, cartoon young woman	Bright colors
Juul	Stylish, intensely satisfying, intelligent design, elegant, innovative, young people	Bright colors, design
ExcluCig	Exclusive, luxurious, fashionable, high quality, young woman	—[a]
Treasurer vape	Elegant, discrete, pure, high quality, high-end product	White, light grey, flowers, design
Vaporcade Jupiter	Technology, discrete, quality, young woman	Black, design
Innokin lily	Elegant, luxurious, exclusive, beautiful vaping, young woman, highest quality, design	Swarovski crystals, flower, colors
Zensations	Unique, like real cigarette, variation of tastes	Design
Cig-a-LinQ	A Dutch brand, next generation, developed by smokers	Stylish

[a] Information not available.

Table 2. Packaging and labeling information of e-liquids upon online ordering from several Dutch Web shops. Information as visibility of a health warning on the package or on the website as well as nicotine content and age verification are reported. A total number of 25 Web shops were randomly selected; from each Web shop, at least 2 products were assessed. Information was obtained before the Tobacco Products Directive came into force.

Information on packaging or website	Not present (%)	Present (%)
Health warning visible on packaging	84	16
Health warning on website	80	20
Nicotine content indicated	0	100
Age verification	60	40

Design and Packaging

The latest type of customizable e-cigarettes with Bluetooth function, colored led lights, and vitamins show that not just the technology is changing toward the user's needs but also their functionality and appearance.

Stylish packaging, product design, and color choices of the product, such as stainless steel, match marketing terms such as "elegance' and 'discrete." Moreover, the flavors, colors, naming, and electronic capabilities of the products could specifically appeal to young people and unexperienced users. Flavors that are added to e-liquids are often sweet and fruity (eg, apple, melon) or tobacco-like or they give the impression of gaining energy (coffee, energy) or a soothing sense of holiday spirit or relaxation (cocktails, tropical, cannabis). A significant number of Europeans consider a reference to flavors and pack color to be indicative for the level of harm of a tobacco product [35].

Health Warnings

Consumers consider health warnings and nicotine levels presented on a product to indicate the level of harm [35]. We assessed the packaging and labeling of 25 randomly selected e-liquids available for sale on Web shops in the Netherlands (Table 2). We found that nicotine content was often indicated in the advertisement text, but health warnings were generally not visible on product photos of the outer package nor on the website of the shop selling them. This is in contradiction to packages of regular cigarettes, almost all of which display a visible health warning on the package. Verification of age of the buyer and information on health effects of the product upon purchase were not clearly present or even absent when buying either tobacco or electronic cigarette products online. In only a quarter of the visited Web shops, an age verification was requested to gain access (Table 2), which can still easily be circumvented by people under 18 years.

Public Interest

We analyzed the number of searches on Google to have an indication of the regional interest in different types of e-cigarettes and the changes over time. Public interest in different types of e-cigarettes, such as eGo-, Evod-, and Evic series, are visualized over the past few years in Figures 1 and 2.

These figures show that the number of searches, as an indicator of popularity of a new product, could increase rapidly, but also decline rapidly. The number of searches on products that are available on the market is relatively stable after a while.

Cigarette-Like Products That Heat Tobacco

Trends in Products That Heat Tobacco

Cigarettes that heat tobacco were reintroduced on the market in 2014 by Reynolds (Revo) and Phillip Morris (iQOS). The Revo, by Reynolds, is a cigarette whereby heating is performed using a carbon tip wrapped in glass fibers, which was marketed earlier under the name "Eclipse." Phillip Morris replaced their formed heatbar (introduced in 2007) by the iQOS, which is a type of electronic cigarette that can heat tobacco. The tobacco product to be heated was to be marketed under the Marlboro brand (currently known as Marlboro heets). Another heat-not-burn device is the Ploom, developed by the Ploom Company and taken over by Japan Tobacco Inc. The Ploom Company, now called PAX Labs Inc, also developed the Pax and the Pax 2. These are both products that vaporize heated tobacco. Furthermore, The Firefly developed The Firefly 2, which heats loose-leaf plant material and concentrates and is often used to vaporize marihuana. During the time of our search, British American Tobacco was developing a product that combines tobacco and e-cigarette technology by heating nicotine-laced liquid into an inhalable vapor that passes through a bit of tobacco near the tip (iFUSE) [36].

Marketing, Packaging, and Labeling

Terms used in marketing of cigarette-like products that "heat rather than burn" are referring to the product as "reduced risk" and "innovative." The appearance of these products is modern and elegant (Table 3).

Figure 1. Searches on specific types of e-cigarettes worldwide over time (January 2012 – February 2016) using Google Trends. The interest (y-axis) is shown relative to the total number of searches on Google in this time window.

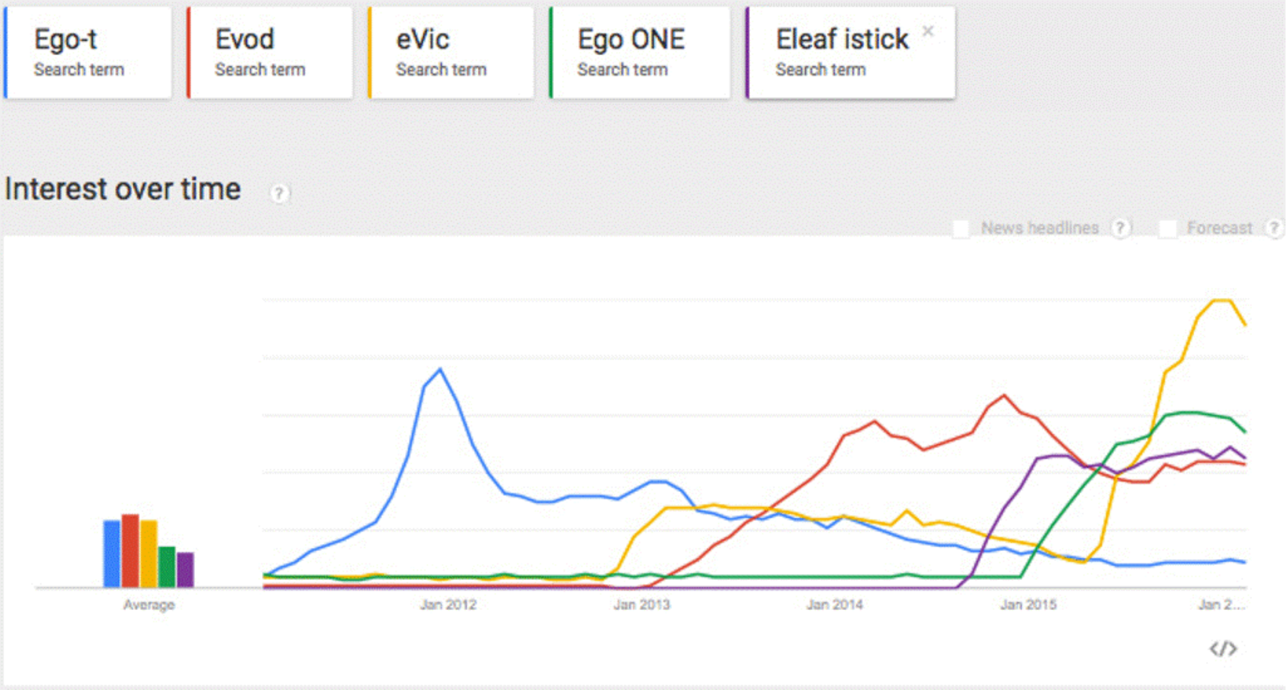


Figure 2. Searches on specific types of e-cigarettes in the Netherlands (1b) over time (January 2012-February 2016) using Google Trends. The interest (y-axis) is shown relative to the total number of searches on Google in this time window.

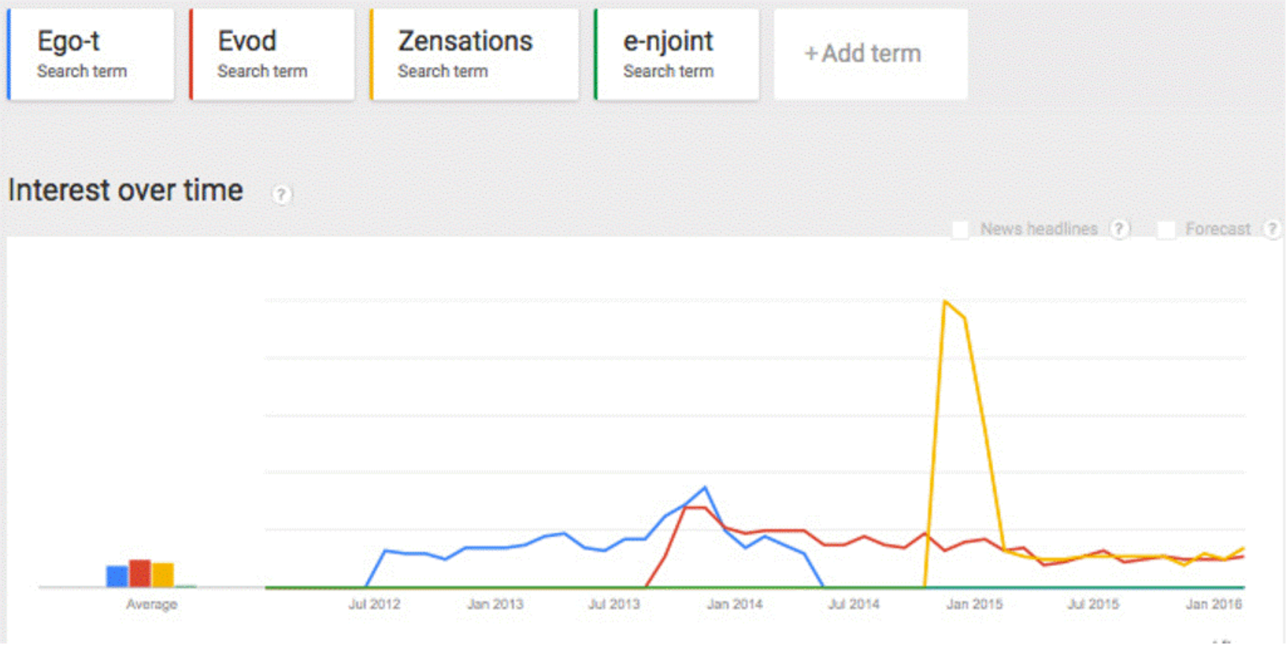
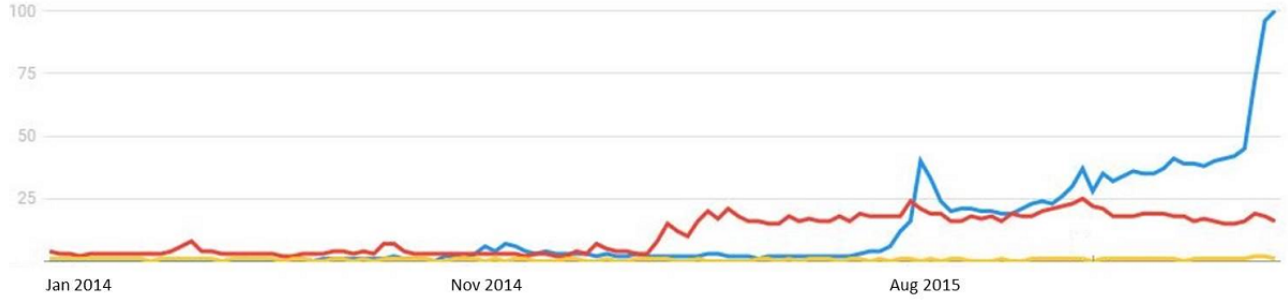


Table 3. Product appeal and marketing terms used for online advertisement of cigarette-like products.

Product name	Marketing terms	Product appeal
iQOS	Reduced risk product, innovative	Clean (white, bright blue), stylish, elegant
Revo	Reduced risk	Similar sized package as traditional cigarette, white or light grey or gold
PAX 2	Smaller, smarter, sleeker	Design, elegant and fun
iFuse	Reduced risk	Packed as traditional cigarette, stylish

Figure 3. Relative number of searches for heat-not-burn products: iQOS in blue, Pax 2 in red and iFUSE in yellow, worldwide in the period January 2014- May 2016) using Google Trends.



Public Interest

Figure 3 shows the relative frequency of internet searches for heat-not-burn products. The iQOS was first released in November 2014 in 2 pilot projects conducted in the cities of Nagoya (Japan) and Milan (Italy) [37]. In September 2015, the availability of the iQOS was expanded in Japan and was introduced in Switzerland. Analysis of worldwide searches on product information only peaks in the 2 countries of market release: Japan and Italy. Along with its market introduction, interest in the product from Swiss citizens was also detected. Most internet searches were performed over time in Japan, indicating the product is gaining popularity there. Most searches on the Pax 2 are conducted in the United States and Canada and seem quite stable over time. In general, and as expected, the relative number of searches and the location of searches coincide with the availability of the product.

Discussion

Principal Findings

This paper describes our search for tobacco(-related) products, and, in retrospect, the qualitative data on trends we found over the period January 2014-May 2016. By means of our search strategy, we were able to identify new developments and trends in tobacco-related products and provide information on new products. For example, we identified a trend toward e-cigarette devices that can be more and more adjusted to personal wishes and needs. This was evident from the many available accessories that can be used to adjust physical properties of the e-cigarette, such as airflow. Moreover, our search identified some e-cigarette-like devices with very specific functions, such as playing music or making calls. Additionally, mobile phone applications are available that connect to an e-cigarette and keep track of user behavior. Besides e-cigarettes, our search found that several devices were introduced to the market that heat tobacco instead of burning it. We also found waterpipe products in our search, but we did not elaborate on those in our paper.

Most of the identified new products were marketed using terms indicating less harm and an elegant or luxurious lifestyle. This is in line with an earlier study by Escobedo et al [38], who found that e-cigarette websites were much more likely to feature themes related to harm reduction. In addition, marketing of some products was aimed at specific target groups such as young/unexperienced consumers or smokers who may switch to these new products. Although advertisements or promotions for tobacco products, including e-cigarettes, are currently no longer permitted, manufacturers can still make their products more attractive (for certain groups) by changing their design and appearance. In line with this, the designs of the products in our search were characterized as stylish and modern.

Although we were able to identify products that seemed to be gaining popularity during the time scope of our search, we cannot draw conclusions about the actual use of the products based on our data. That is, the popularity of products in our study was inferred using the trend data from social media, which reflect the public interest in a specific product rather than the actual use of it [39]. However, as this approach has been taken before [26] and provides an indication of the consumer perception [18], this seems to provide a valid estimate of popularity. Although, as e-cigarette proponents appear to be overrepresented on social media [40], this may have biased our findings of the Google trend analysis. We suggest that, when a product appears to increase in popularity based on social media trends, information from other sources (eg, sales data and prevalence data from monitoring studies) should be consulted to confirm whether there is also an actual increase in use.

A limitation of this study is that our approach to our analysis of websites and online retailers could have been more systematic. This study started out as an exploratory search for new products, aiming to find as much information on new products as possible from all thinkable sources. A more structured, systematic approach would have been useful for selecting and classifying relevant information in a clear and objective manner. Previous studies have employed machine learning approaches to classify data from social media [12]. Such approaches seem also suitable for the identification of novel tobacco-related products, as Allem at al identified 2 new e-cigarette devices by means of a social network analyses of Twitter data. Classification of social media data has not only been shown to be valuable in identification of new products but can also be used for surveillance of product use [12]. Moreover, it can be used to identify and classify the sources of social media messages (eg, retailers, product users), to gain more insight in marketing strategies [41]. Related to this, manufacturers and retailers use social media to increase the popularity of a product such as the e-cigarette, for instance by posting specific types of images [42]. They also promote combined use of products, for instance the use of waterpipe with alcoholic beverages [43]. In addition, other platforms have been used to promote e-cigarettes, such as the Pokémon Go game [16]. Alternatively, social media is also used to promote healthier behavior [44], which can influence the public's opinion on specific tobacco-related products. For instance, the products that were identified in our search were marketed using terms indicating less harm. Systematic analyses of social media data could thus provide valuable insight in current marketing strategies, which in their turn can influence attractiveness and popularity of specific products.

Overall, our search strategy was suitable for our aims, as it identified some interesting developments. It was also specific, as we received little or no messages outside our scope. By verification with messages from other sources, the completeness of the search tool can be checked. Other sources are, for example, information received from newsletters, information from colleagues in the field, and manual internet searches. In addition, search terms or websites were added to optimize the automated search. The data obtained from the search were also timely, as the time between a notification by our search tool and the actual market introduction was short. For example, the introduction of the E-njoint on the Dutch market was identified in our search a few days before it was officially reported by the Netherlands Institute of Mental Health and Addiction (Trimbos Instituut).

Implications for Policy and Research

The main concerns related to the use of new tobacco products include unknown toxicity, changes in product use behavior, decreased cessation, increased initiation, sustained prevalence of tobacco "dual use," and public misunderstanding about the actual risk associated with allegedly less hazardous products. In the United States, the Food and Drug Administration is notified of new products or product changes on the US market [45]. Currently, in the European Union, producers are obliged to notify the national government of their intention to market a new product at least 6 months before market introduction. General guidelines for assessing the risks associated with modified tobacco products have been proposed by the Tobacco Product Scientific Advisory Committee and by the Society for Research on Nicotine and Tobacco [6,46]. This notification allows preparation of requirements, if needed, for communication and marketing of such products to avoid misinterpretation of the harmful effects of the product for the general public [47]. As this information is only available nationally and Web shops facilitate the availability of products almost worldwide, it is important to follow developments worldwide. The information from the producer, together with the information from search strategies like ours, can be used to follow a product over time and may be used for law enforcement. The information can be used to assess the likelihood of new products to gain market share, either as a possible health risk for the general population or as an indication for a need for independent and reliable information of such products to the general public.

Acknowledgments

The authors thank Rob van Spronsen for setting up the algorithm for the literature and news search and regularly providing data from these searches.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Keywords.

[PDF File (Adobe PDF File), 20KB - publichealth_v4i2e55_app1.pdf]

Multimedia Appendix 2

Websites searched for the selected keywords.

[PDF File (Adobe PDF File), 26KB - publichealth_v4i2e55_app2.pdf]

References

1. Kienhuis AS, Soeteman-Hernandez LG, Bos PM, Cremers HW, Klerx WN, Talhout R. Potential harmful health effects of inhaling nicotine-free shisha-pen vapor: a chemical risk assessment of the main components propylene glycol and glycerol. Tob Induc Dis 2015;13(1):15 [FREE Full text] [doi: 10.1186/s12971-015-0038-7] [Medline: 26120296]
2. Agaku IT, Omaduvie UT, Filippidis FT, Vardavas CI. Cigarette design and marketing features are associated with increased smoking susceptibility and perception of reduced harm among smokers in 27 EU countries. Tob Control 2015 Dec;24(e4):e233-e240. [doi: 10.1136/tobaccocontrol-2014-051922] [Medline: 25335899]
3. Slater S, Giovino G, Chaloupka F. Surveillance of tobacco industry retail marketing activities of reduced harm products. Nicotine Tob Res 2008 Jan;10(1):187-193. [doi: 10.1080/14622200701704947] [Medline: 18188759]
4. Staal Y, Talhout R. Rivm.nl. 2016. Alternatieve tabaksproducten: harm reduction? Tabaks- en aanverwante producten die mogelijk minder schadelijk zijn dan sigaretten URL: http://www.rivm.nl/dsresource?objectid=79d947d2-a420-489e-8bfa-e8d7bf7dcf77&type=org&disposition=inline:
5. Stepanov I, Soeteman-Hernandez LG, Talhout R. Research and monitoring the evolution of new tobacco products, including products with potentially “modified risks”. 2013 Presented at: Background Paper for 7th Meeting of WHO TobReg; 2013; Rio de Janeiro, Brazil p. 2013.
6. World Health Organization. 2014. Work in Progress in Relation to Articles 9 and 10 of the WHO FCTC URL: http://apps.who.int/gb/fctc/PDF/cop6/FCTC_COP6_14-en.pdf [WebCite Cache ID 6z5lfYAvA]
7. Czoli CD, Goniewicz M, Islam T, Kotnowski K, Hammond D. Consumer preferences for electronic cigarettes: results from a discrete choice experiment. Tob Control 2015 Oct;25(e1):e30-e36. [doi: 10.1136/tobaccocontrol-2015-052422] [Medline: 26490845]
8. Wackowski OA, Hammond D, O'Connor RJ, Strasser AA, Delnevo CD. Smokers' and e-cigarette users' perceptions about e-cigarette warning statements. Int J Environ Res Public Health 2016 Jun 30;13(7) [FREE Full text] [doi: 10.3390/ijerph13070655] [Medline: 27376310]
9. Jawad M, Bakir A, Ali M, Grant A. Impact of waterpipe tobacco pack health warnings on waterpipe smoking attitudes: a qualitative analysis among regular users in London. Biomed Res Int 2015;2015:745865 [FREE Full text] [doi: 10.1155/2015/745865] [Medline: 26273642]
10. Li J, Newcombe R, Walton D. The use of, and attitudes towards, electronic cigarettes and self-reported exposure to advertising and the product in general. Aust N Z J Public Health 2014;38(6):524-528. [doi: 10.1111/1753-6405.12283] [Medline: 25376856]
11. Lazard AJ, Saffer AJ, Wilcox GB, Chung AD, Mackert MS, Bernhardt JM. E-cigarette social media messages: a text mining analysis of marketing and consumer conversations on Twitter. JMIR Public Health Surveill 2016 Dec 12;2(2):e171 [FREE Full text] [doi: 10.2196/publichealth.6551] [Medline: 27956376]
12. Allem JP, Ferrara E, Uppu SP, Cruz TB, Unger JB. E-cigarette surveillance with social media data: social bots, emerging topics, and trends. JMIR Public Health Surveill 2017 Dec 20;3(4):e98 [FREE Full text] [doi: 10.2196/publichealth.8641] [Medline: 29263018]
13. Williams RS, Derrick J, Liebman AK, LaFleur K. Content analysis of e-cigarette products, promotions, prices and claims on Internet tobacco vendor websites, 2013-2014. Tob Control 2017 Nov 03. [doi: 10.1136/tobaccocontrol-2017-053762] [Medline: 29101294]
14. Caputi TL, Leas E, Dredze M, Cohen JE, Ayers JW. They're heating up: internet search query trends reveal significant public interest in heat-not-burn tobacco products. PLoS One 2017;12(10):e0185735 [FREE Full text] [doi: 10.1371/journal.pone.0185735] [Medline: 29020019]
15. Allem JP, Ayers JW, Althouse BM, Williams R. When a ban really is not a ban: internet loopholes and djarum flavoured cigarettes in the USA. Tob Control 2016;25(4):489-490 [FREE Full text] [doi: 10.1136/tobaccocontrol-2015-052309] [Medline: 26085123]
16. Kirkpatrick MG, Cruz TB, Goldenson NI, Allem JP, Chu KH, Pentz MA, et al. Electronic cigarette retailers use Pokémon Go to market products. Tob Control 2017;26(e2):e145-e147. [doi: 10.1136/tobaccocontrol-2016-053369] [Medline: 28044010]
17. Rose SW, Jo CL, Binns S, Buenger M, Emery S, Ribisl KM. Perceptions of menthol cigarettes among Twitter users: content and sentiment analysis. J Med Internet Res 2017 Feb 27;19(2):e56 [FREE Full text] [doi: 10.2196/jmir.5694] [Medline: 28242592]
18. Ayers JW, Althouse BM, Allem JP, Leas EC, Dredze M, Williams RS. Revisiting the rise of electronic nicotine delivery systems using search query surveillance. Am J Prev Med 2016 Jun;50(6):e173-e181. [doi: 10.1016/j.amepre.2015.12.008] [Medline: 26876772]
19. Vaporfi. 2016 Jul 25. VaporFi Air Flow Control URL: http://www.vaporfi.com/vape-accessories/vaporfi-air-flow-control.html [accessed 2018-05-01] [WebCite Cache ID 6z5lktFtg]
20. Rahman MA, Hann N, Wilson A, Worrall-Carter L. Electronic cigarettes: patterns of use, health effects, use in smoking cessation and regulatory issues. Tob Induc Dis 2014;12(1):21 [FREE Full text] [doi: 10.1186/1617-9625-12-21] [Medline: 25745382]

21. Want. 2014. Is dit een grapje? Met de Supersmoker Bluetooth kun je roken, bellen en muziek luisteren URL: https://www.want.nl/is-dit-een-grapje-met-de-supersmoker-bluetooth-kun-je-roken-bellen-en-muziek-luisteren/2014/ [WebCite Cache ID 6z5lwLPQB]
22. Daily Mail. 2015. The phone you can SMOKE: Vaporcade Jupiter comes with a built-in e-cigarette to let you 'vape' and dial URL: http://www.dailymail.co.uk/sciencetech/article-3330177/The-phone-SMOKE-299-Vaporcade-Jupiter-comes-built-e-cigarette-let-vape-dial.html?ITO=1490&ns_mchannel=rss&ns_campaign=1490:
23. Digital Trends. 2014. With every puff of vapor, Smokio counts the cigarettes you're not smoking URL: http://www.digitaltrends.com/cool-tech/smokio-first-smart-e-cigarette-tracks-nicotine-consumption/ [WebCite Cache ID 6z5mJT4rR]
24. Trendhunter. 2014. The Future Could Have a WiFi-Connected Smart E-Cigarette URL: https://www.trendhunter.com/trends/smart-ecigarette [WebCite Cache ID 6z5mMwf7f]
25. Vaporcade. 2014. Jupiter IO 4: 4G LTE Cellular Vaping Phone URL: https://store.vaporcade.com/products/jupiter-4g-lte-cellular-vaping-phone [accessed 2018-05-08] [WebCite Cache ID 6zGAe0vWZ]
26. Wang L, Zhan Y, Li Q, Zeng DD, Leischow SJ, Okamoto J. An examination of electronic cigarette content on social media: analysis of e-cigarette flavor content on Reddit. Int J Environ Res Public Health 2015;12(11):14916-14935 [FREE Full text] [doi: 10.3390/ijerph121114916] [Medline: 26610541]
27. Cobb CO, Hendricks PS, Eissenberg T. Electronic cigarettes and nicotine dependence: evolving products, evolving problems. BMC Med 2015 May 21;13:119 [FREE Full text] [doi: 10.1186/s12916-015-0355-y] [Medline: 25998379]
28. Lopez AA, Eissenberg T. Science and the evolving electronic cigarette. Prev Med 2015 Nov;80:101-106 [FREE Full text] [doi: 10.1016/j.ypmed.2015.07.006] [Medline: 26190363]
29. NBC San Diego. 2014. 'E-Hookah' Debuts in San Diego Amid Critics URL: https://www.nbcsandiego.com/news/local/E-Hookah-vape-cigarette-smoke-243042311.html [WebCite Cache ID 6z5mVBgeG]
30. Hookah-shisha. 2016. Hookah Charcoal & Hookah Coal URL: https://www.hookah-shisha.com/c-4-hookah-coals-hookah-charcoal.html2016 [accessed 2018-05-08] [WebCite Cache ID 6zGBVvuyy]
31. Wagoner KG, Cornacchione J, Wiseman KD, Teal R, Moracco KE, Sutfin EL. E-cigarettes, hookah pens and vapes: adolescent and young adult perceptions of electronic nicotine delivery systems. Nicotine Tob Res 2016 Oct;18(10):2006-2012. [doi: 10.1093/ntr/ntw095] [Medline: 27029821]
32. The Verge. 2014. Don't listen to e-cig companies that claim you can vape vitamins URL: https://www.theverge.com/2014/8/28/6078267/don-t-listen-to-e-cig-companies-that-claim-you-can-vape-vitamins [WebCite Cache ID 6z5mdcqdw]
33. Epuff. 2016. Do you want to sweeten your life? URL: http://www.epuff.co.uk/epuff-ambassador/ [accessed 2018-05-01] [WebCite Cache ID 6z5mfZi1r]
34. European Commission. 2014. Directive 2014/40/EU of the European Parliament and of the Council URL: https://ec.europa.eu/health/sites/health/files/tobacco/docs/dir_201440_en.pdf [WebCite Cache ID 6z5mr1sJ5]
35. European Commission. 2015. Attitudes of Europeans Towards Tobacco and Electronic Cigarettes URL: http://ec.europa.eu/public_opinion/archives/ebs/ebs_429_en.pdf [WebCite Cache ID 6z5n00PJO]
36. Forster M, Liu C, Duke MG, McAdam KG, Proctor CJ. An experimental method to study emissions from heated tobacco between 100-200°C. Chem Cent J 2015;9:20 [FREE Full text] [doi: 10.1186/s13065-015-0096-1] [Medline: 25941536]
37. VapeRanks. 2014 Nov 25. Phillip Morris Launches New Type of Smokeless Cigarette URL: http://vaperanks.com/phillip-morris-launches-new-type-of-smokeless-cigarette/25/ [accessed 2018-05-08] [WebCite Cache ID 6zGCYIgCI]
38. Escobedo P, Cruz TB, Tsai KY, Allem JP, Soto DW, Kirkpatrick MG, et al. Monitoring tobacco brand websites to understand marketing strategies aimed at tobacco product users and potential users. Nicotine Tob Res 2017 Sep 11. [doi: 10.1093/ntr/ntx200] [Medline: 29059386]
39. Cole-Lewis H, Pugatch J, Sanders A, Varghese A, Posada S, Yun C, et al. Social listening: a content analysis of e-cigarette discussions on Twitter. J Med Internet Res 2015;17(10):e243 [FREE Full text] [doi: 10.2196/jmir.4969] [Medline: 26508089]
40. Kavuluru R, Sabbir AK. Toward automated e-cigarette surveillance: spotting e-cigarette proponents on Twitter. J Biomed Inform 2016 Jun;61:19-26. [doi: 10.1016/j.jbi.2016.03.006] [Medline: 26975599]
41. Kim A, Miano T, Chew R, Eggers M, Nonnemaker J. Classification of Twitter users who tweet about e-cigarettes. JMIR Public Health Surveill 2017 Sep 26;3(3):e63 [FREE Full text] [doi: 10.2196/publichealth.8060] [Medline: 28951381]
42. Chu KH, Allem JP, Cruz TB, Unger JB. Vaping on Instagram: cloud chasing, hand checks and product placement. Tob Control 2016 Sep 22;26(5):575-578. [doi: 10.1136/tobaccocontrol-2016-053052] [Medline: 27660111]
43. Allem JP, Chu KH, Cruz TB, Unger JB. Waterpipe promotion and use on Instagram: #Hookah. Nicotine Tob Res 2017 Jan 11;19(10):1248-1252. [doi: 10.1093/ntr/ntw329] [Medline: 28077449]
44. Allem JP, Escobedo P, Chu KH, Boley Cruz T, Unger JB. Images of little cigars and cigarillos on Instagram identified by the hashtag #swisher: thematic analysis. J Med Internet Res 2017 Jul 14;19(7):e255 [FREE Full text] [doi: 10.2196/jmir.7634] [Medline: 28710057]
45. Myers ML. Towards more effective FDA premarket review of new tobacco products. Tob Control 2016;25(3):243-244. [doi: 10.1136/tobaccocontrol-2016-053055] [Medline: 27068244]
46. World Health Organization. 2015. WHO Study Group on Tobacco Product Regulation URL: http://apps.who.int/iris/bitstream/10665/161512/1/9789241209892.pdf [WebCite Cache ID 6z5n8zKe7]

47. Boynton MH, Agans RP, Bowling JM, Brewer NT, Sutfin EL, Goldstein AO, et al. Understanding how perceptions of tobacco constituents and the FDA relate to effective and credible tobacco risk messaging: a national phone survey of U.S. adults, 2014-2015. BMC Public Health 2016 Jun 23;16:516 [FREE Full text] [doi: 10.1186/s12889-016-3151-5] [Medline: 27333921]

Edited by G Eysenbach; submitted 20.01.17; peer-reviewed by B Majeed, JP Allem; comments to author 26.02.17; revised version received 30.03.17; accepted 26.02.18; published 28.05.18.

Please cite as:
Staal YCM, van de Nobelen S, Havermans A, Talhout R
New Tobacco and Tobacco-Related Products: Early Detection of Product Development, Marketing Strategies, and Consumer Interest
JMIR Public Health Surveill 2018;4(2):e55
URL: http://publichealth.jmir.org/2018/2/e55/
doi:10.2196/publichealth.7359
PMID:

Original Paper

Contents, Followers, and Retweets of the Centers for Disease Control and Prevention's Office of Advanced Molecular Detection (@CDC_AMD) Twitter Profile: Cross-Sectional Study

Isaac Chun-Hai Fung[1*], MSc, PhD; Ashley M Jackson[1*], MPH; Lindsay A Mullican[1*], MPH; Elizabeth B Blankenship[1], MPH; Mary Elizabeth Goff[1], MPH; Amy J Guinn[2], MA; Nitin Saroha[3], BS; Zion Tsz Ho Tse[4], PhD

[1]Department of Epidemiology and Environmental Health Sciences, Jiann-Ping Hsu College of Public Health, Georgia Southern University, Statesboro, GA, United States

[2]Office of Advanced Molecular Detection, National Center for Emerging and Zoonotic Infectious Diseases, Centers for Disease Control and Prevention, Atlanta, GA, United States

[3]Department of Computer Science, The University of Georgia, Athens, GA, United States

[4]School of Electrical and Computer Engineering, College of Engineering, The University of Georgia, Athens, GA, United States

[*]these authors contributed equally

Corresponding Author:
Isaac Chun-Hai Fung, MSc, PhD
Department of Epidemiology and Environmental Health Sciences
Jiann-Ping Hsu College of Public Health
Georgia Southern University
PO Box 8015
Statesboro, GA, 30460
United States
Phone: 1 9124785079
Fax: 1 9124780171
Email: cfung@georgiasouthern.edu

Abstract

Background: The Office of Advanced Molecular Detection (OAMD), Centers for Disease Control and Prevention (CDC), manages a Twitter profile (@CDC_AMD). To our knowledge, no prior study has analyzed a CDC Twitter handle's entire contents and all followers.

Objective: This study aimed to describe the contents and followers of the Twitter profile @CDC_AMD and to assess if attaching photos or videos to tweets posted by @CDC_AMD would increase retweet frequency.

Methods: Data of @CDC_AMD were retrieved on November 21, 2016. All followers (N=809) were manually categorized. All tweets (N=768) were manually coded for contents and whether photos or videos were attached. Retweet count for each tweet was recorded. Negative binomial regression models were applied to both the original and the retweet corpora.

Results: Among the 809 followers, 26.0% (210/809) were individual health professionals, 11.6% (94/809) nongovernmental organizations, 3.3% (27/809) government agencies' accounts, 3.3% (27/809) accounts of media organizations and journalists, and 0.9% (7/809) academic journals, with 54.9% (444/809) categorized as miscellaneous. A total of 46.9% (360/768) of @CDC_AMD's tweets referred to the Office's website and their current research; 17.6% (135/768) referred to their scientists' publications. Moreover, 80% (69/86) of tweets retweeted by @CDC_AMD fell into the miscellaneous category. In addition, 43.4% (333/768) of the tweets contained photos or videos, whereas the remaining 56.6% (435/768) did not. Attaching photos or videos to original @CDC_AMD tweets increases the number of retweets by 37% (probability ratio=1.37, 95% CI 1.13-1.67, *P*=.002). Content topics did not explain or modify this association.

Conclusions: This study confirms CDC health communicators' experience that original tweets created by @CDC_AMD Twitter profile sharing images or videos (or their links) received more retweets. The current policy of attaching images to tweets should be encouraged.

(JMIR Public Health Surveill 2018;4(2):e33) doi:10.2196/publichealth.8737

KEYWORDS

communications media; health communication; social media

Introduction

Twitter (San Francisco, CA) has been used by public health practitioners for purposes ranging from public health surveillance [1] to health communication [2,3] to natural disaster preparedness [4]. A recent systematic review identified taxonomy of Twitter as a tool for health research, including content analysis, surveillance, engagement, recruitment, intervention, and network analysis [5]. Prior research found Twitter users engaged in communication pertinent to infectious diseases, such as Ebola [6,7], measles [8], Middle East Respiratory Syndrome [9], and Zika [10,11]. The Centers for Disease Control and Prevention (CDC) encourages the strategic use of Twitter to disseminate CDC health information and engage with individuals and partners [12]. CDC uses Twitter as part of their overall health communication strategies [13,14], uses Twitter chats to engage Twitter users on specific health topics [15-17], and publicizes public health events and publications via Twitter [18].

Advanced molecular detection (AMD) harnesses the power of next-generation genomic sequencing, high-performance computing, and epidemiology to study pathogens. CDC uses AMD technologies to identify emerging pathogens, improve vaccines, make food safer, develop faster tests, and connect information from public health investigations with genomic data from pathogens to understand how infections spread [19].

The Office of Advanced Molecular Detection (OAMD) in the National Center for Emerging and Zoonotic Infectious Diseases (NCEZID), CDC, promotes open sharing of scientific data through publicly accessible platforms and added Twitter (@CDC_AMD) as a communication channel in May 2015. OAMD uses Twitter to promote free exchange of information and interactivity between CDC infectious disease programs and partners in federal and state agencies, academia, and professional organizations, as well as the general public. OAMD noted the value of Twitter in March 2016, when a tweet regarding CDC's submission of genomic sequence data to a publicly accessible database led to a global discussion on *Elizabethkingia anophelis* and a subsequent international partnership to investigate this rare bacterium [20].

In this case study, we analyzed the @CDC_AMD Twitter handle to divulge information regarding followers and popular tweet content. Here, we address 3 research questions (RQ):

> RQ1: Who are the followers of @CDC_AMD?
>
> RQ2: What content has been tweeted the most?
>
> RQ3: Does attaching a photo or video (or a link to a photo or video) increase the probability ratio of a tweet being retweeted?

This case study enables us to understand how a CDC Twitter account communicates scientific information to its followers, and provides health communicators with information for future enhancement of their Twitter communication.

Methods

Data

Data were retrieved via Twitter Search Application Programming Interface and downloaded to a server at Athens, Georgia, USA. Tweets tweeted by OAMD from the first tweet on May 5, 2015, 3:04 pm (Universal Coordinated Time, UTC) to the 768th tweet on November 16, 2016, 3:22 pm UTC were retrieved. Data on favorites, followers, followings, and tweets, including retweet count, were downloaded by November 21, 2016. A protocol for analysis was created and approved for use by OAMD.

CDC encourages the strategic use of Twitter to disseminate CDC health information and engage with individuals and partners. OAMD's Twitter guidelines, based on CDC's Twitter Guidelines and Best Practices [12], give careful consideration to the nature of @CDC_AMD Twitter messages and activities. All tweets posted by @CDC_AMD are cleared internally by OAMD scientific and communication staff or cross-cleared with scientific and communication staff in other CDC programs that are conducting AMD-related activities.

Manual Coding

First, we manually categorized @CDC_AMD's followers into 6 categories specified by OAMD:

1. Individual scientists, physicians, and other public health professionals
2. Governmental organizations, such as other branches in CDC, and other federal agencies
3. Nongovernmental organizations, such as a scientific society
4. Mass media organizations and their representatives, such as Cable News Network, and individual journalists
5. Academic journals, such as *Journal of the American Medical Association* and *Science*
6. Miscellaneous: anyone who did not belong to the aforementioned categories

Next, we manually categorized @CDC_AMD's tweets into 6 categories specified by OAMD:

1. Tweets that refer to CDC's AMD website [21]
2. Tweets that refer to publications of CDC's AMD scientists (usually their abstract on PubMed)
3. Training: announcement of webinars, every quarter, in collaboration with Association of Public Health Laboratories (APHL)
4. Training: announcement of CDC Bioinformatics fellowship program, in collaboration with APHL
5. CDC AMD scientists' activities, such as their visit to a state laboratory
6. Miscellaneous: anything that does not belong to the aforementioned categories

Then, we manually determined whether photos or videos, or their links, were attached to the tweets. We also manually coded whether a tweet was an original tweet posted by @CDC_AMD or it was a retweet by @CDC_AMD of a tweet originally posted by other Twitter users.

A total of 10% random samples of original tweets posted by @CDC_AMD (68/682), @CDC_AMD's retweets of other Twitter users' tweets (9/86), and followers of @CDC_AMD (81/809) were double-coded by 2 independent coders, and the inter-rater reliability is substantial (κ=.749, .757, and .839, respectively) [22].

Statistical Analysis

Statistical analysis was conducted in R 3.3.1 (R Core Team, R Foundation for Statistical Computing, Vienna, Austria) via RStudio 0.99.903 (RStudio, Inc., Boston, MA). For Poisson regression models, we used glm() in stats package. For negative binomial regression models, we used glm.nb() in MASS package. For hurdle models, we used hurdle() in pscl package [23].

We stratified the corpus of tweets into 2 subcorpora (original tweets and tweets of other Twitter users retweeted by @CDC_AMD). Regression models were then applied to each subcorpus to compute the probability ratios, to evaluate whether sharing images or videos (or their links) increased the probability of a tweet being retweeted, and whether contents might be a confounder or an effect modifier. Negative binomial regression models were used, as the retweet frequency was overdispersed. Negative binomial regression models took into account data points that were zeros. Hurdle models that take into account the excess of zeros (because many tweets did not have any retweet) were also attempted. A hurdle model is a model with 2 components: (1) a hurdle component for zero versus positive counts and (2) a truncated count component for positive counts [23]. In our hurdle models, the default binomial distribution (logistic regression) was used for the *hurdle component*; a truncated negative binomial distribution was chosen for the *count component*, given the overdispersion of the data.

Original Tweets by @CDC_AMD

For the original tweet subcorpus, we performed model selection between 5 different statistical models. We started with a Poisson regression model with both the variable for photo or video attachment (*Media*) and the content variable (*Content*) as predictor variables and retweet count as the outcome variable (Model A). Due to overdispersion of the retweet count data, we explored the use of negative binomial regression models (Models B and C) and the hurdle models (Models D and E), and we explored whether content is a confounder by comparing models that include both Media and Content (Models B and D) and those that only include the Media predictor (Models C and E), through likelihood ratio tests and Akaike Information Criterion (AIC).

Retweets by @CDC_AMD

We constructed 2 negative binomial regression models. Model F included both Media and Content variables as predictors; Model G included Media only.

Ethics Approval

This study was approved (H15083) by Georgia Southern University's Institutional Review Board under the B2 exempt category because of the fact that the social media posts analyzed in this study are considered publicly observable behavior.

Availability of Data and Material

The datasets used and/or analyzed during this study are available from the corresponding author on reasonable request.

Results

Findings

As of November 21, 2016, @CDC_AMD had published 768 tweets, of which 88.8% (682/768) were original and 11.2% (86/768) were @CDC_AMD's retweets of other Twitter users' tweets. @CDC_AMD had 809 followers, and it followed 195 Twitter accounts. There were 37 likes.

Followers

Among the 809 followers, 26.0% (210/809) were individual health professionals, 11.6% (94/809) nongovernmental organizations, 3.3% (27/809) government agencies' accounts, 3.3% (27/809) accounts of media organizations and journalists, and 0.9% (7/809) academic journals. In addition, 54.9% (444/809) of the followers did not fall under any prespecified categories.

Content

Approximately half of @CDC_AMD's tweets (360/768, 46.9%) referred to the CDC AMD website and research that was taking place during that time; of these tweets, the majority (354/360, 98.3%) were original tweets (Table 1). Nearly 1 in 5 (135/768, 17.6%) tweets referred to publications of CDC AMD scientists (usually their abstracts on PubMed), of which all, except 2 (133/135, 98.5%), were original tweets. Moreover, 8 in 10 (69/86, 80%) tweets retweeted by @CDC_AMD fell into the miscellaneous category.

Tweets by @CDC_AMD that fell into the miscellaneous category related to a variety of scientific activities that were not included in the prespecified categories. For example, 1 tweet promoted an exhibition in the CDC Museum: "Join us this Thu, Feb 4, 6-8 pm, to see AMD in action in @CDC Museum's opening of Places & Spaces: Mapping Science. https://t.co/Wp3eZVleUe." (February 2, 2016; Tweet ID: 694581586566717440); another highlighted the program's annual science event, AMD Day: "Thanks to @APHL state partners and all who attended #AMDDay2016. " (September 26, 2016; Tweet ID: 780511838639448064).

Tweets retweeted by @CDC_AMD that fell into the miscellaneous category covered various scientific topics tweeted by other CDC Twitter handles such as @CDCgov, @DrFriedenCDC, and @CDC_NCEZID, as well as tweets posted by other scientific organizations. For example, the most retweeted tweet in our dataset of 768 tweets was about the Google Doodle on scientist Antoine van Leeuwenhoek. On October 24, 2016, @CDC_AMD retweeted a tweet tweeted by the American Society of Microbiology, "RT @ASMicrobiology: Proud to see today's Google Doodle commemorating the birth of one of the founders of our science Antoine van Leeuwenh..." (The tweet ID of the original tweet by @ASMicrobiology:

790551432734842880; the retweet by @CDC_AMD: 790618487421042688). Regarding photos or videos attached to tweets, 43% (333/768) of the tweets contained photos or videos, whereas the remaining 57% (435/768) did not.

Retweets

For the subcorpus with original tweets, the results of our model selection process are presented in Table 2. A likelihood ratio test between Model A (Poisson regression) and Model B (negative binomial regression) gave a significant result (χ^2_1=259.7, *P*<.001), indicating that Model B, with a higher log-likelihood, fit the data better. A likelihood ratio test between Model B (negative binomial regression with "Content" as confounder) and Model C (negative binomial regression without "Content") found that there was no significant difference in their log-likelihood (χ^2_5=9.3, *P*=.10). The simple model (Model C) had a lower AIC score and was therefore preferred.

Given the excess of zeros in the data, we explored the hurdle models as aforementioned. A likelihood ratio test between Model D (hurdle model with "Content" as confounder; Table 3) and Model E (hurdle model without "Content"; Table 4) found that there was no significant difference in their log-likelihood (χ^2_{10}=17.9, *P*=.06). With a lower AIC score, Model E was therefore preferred (Table 2). Finally, a likelihood ratio test between Model C (negative binomial model) and Model E (hurdle model) found that there was no significant difference in their log-likelihood (χ^2_2=0.7, *P*=.70). As Model C has a lower AIC, we chose Model C as the final model for the subcorpus of original Twitter content posted by @CDC_AMD (Table 2).

An original tweet from @CDC_AMD sharing images or videos (or their links) had 37% more retweets (Model C: probability ratio=1.374, 95% CI 1.129-1.674, *P*=.002) than that of an original tweet that did not share images or videos (or their links; Table 3). We observed no significant difference between the content categories, except for the miscellaneous category that has 36% more retweets than tweets referred to the CDC AMD website (Model B: probability ratio=1.355, 95% CI 1.035-1.778, *P*=.03; Model D: probability ratio=1.501, 95% CI 1.046-2.153, *P*=.03).

Table 1. Frequency of tweets by their content category.

Content category	Original tweets by @CDC_AMD, n (%)	Retweets of other Twitter users' tweets by @CDC_AMD, n (%)	All tweets posted by @CDC_AMD, n (%)
Tweets that refer to the CDC[a] AMD[b] website	354 (51.9)	6 (7)	360 (46.9)
Tweets that refer to publications of CDC AMD scientists (usually their abstracts on PubMed)	133 (19.5)	2 (2)	135 (17.6)
Training: announcement of webinars, every quarter, collaborated with APHL[c]	26 (3.8)	2 (2)	28 (3.6)
Training: announcement of CDC Bioinformatics fellowship program, collaborated with APHL	29 (4.3)	4 (5)	33 (4.3)
CDC AMD scientists' activities, such as their visit to a state	15 (2.2)	3 (3)	18 (2.3)
Miscellaneous: anything that does not belong to the aforementioned categories	125 (18.3)	69 (80)	194 (25.3)
Total	682 (100.0)	86 (100)	768 (100.0)

[a] CDC: Centers for Disease Control and Prevention.

[b] AMD: advanced molecular detection.

[c] APHL: Association of Public Health Laboratories.

Table 2. Number of parameters, log-likelihood, and Akaike Information Criterion for the 5 models that we tested for the corpus of original tweets created by the @CDC_AMD Twitter profile.

Model	A	B	C	D	E
Model choice	Poisson	Negative binomial	Negative binomial	Hurdle[a]	Hurdle[a]
Predictors[b]	Media+Content	Media+Content	Media only	Media+Content	Media only
Number of parameters	8	9	4	16	6
Log-likelihood (df[c])	−1210.125 (df=7)	−1080.27 (df=8)	−1084.925 (df=3)	−1075.625 (df=15)	−1084.567 (df=5)
Akaike Information Criterion	2434.25	2176.5	2175.9	2181.25	2179.133

[a] Hurdle model (count=negative binomial; zero hurdle=logistic).

[b] Media: attachment of a photo or a video (or a link to a photo or a video).

[c] df: degrees of freedom.

Table 3. Probability ratios for an original @CDC_AMD tweet being retweeted in a negative binomial regression model that includes both the variable for photo or video attachment and the content variable (Model B), and one without content variable (Model C); in a hurdle model that includes both the variable for photo or video attachment and the content variable (Model D) and one without the content variable (Model E).

Explanatory variables of each model	Probability ratio (95% CI)	*P* value
Model B (negative binomial model)		
Contained a photo or video	1.406 (1.114-1.775)	.004
Content		
Referred to the CDC[a] AMD[b] website	Reference	-
Publication in content	0.874 (0.638-1.197)	.40
Webinar in content	1.081 (0.643-1.832)	.77
Bioinformatics in content	1.381 (0.834-2.311)	.21
Scientist in content	1.066 (0.544-2.120)	.85
Miscellaneous content	1.355 (1.035-1.778)	.03
Model C (negative binomial model)		
Contained a photo or video	1.374 (1.129-1.674)	.002
Model D (hurdle model)[c]		
Zero or positive (logistic): Contained a photo or video	1.643 (1.137-2.374)	.008
Zero or positive (logistic): Content		
Referred to the CDC AMD website	Reference	-
Publication in content	1.179 (0.740-1.878)	.49
Webinar in content	1.329 (0.579-3.046)	.50
Bioinformatics in content	2.260 (0.965-5.296)	.06
Scientist in content	1.685 (0.557-5.097)	.36
Miscellaneous content	1.150 (0.743-1.782)	.53
Positive count (negative binomial): Contained a photo or video	1.257 (0.917-1.724)	.16
Positive count (negative binomial): Content		
Referred to the CDC AMD website	Reference	-
Publication in content	0.652 (0.417-1.022)	.06
Webinar in content	0.929 (0.463-1.867)	.84
Bioinformatics in content	0.999 (0.511-1.950)	>.99
Scientist in content	0.774 (0.314-1.907)	.58
Miscellaneous content	1.501 (1.046-2.153)	.03
Model E (Hurdle model)[c]		
Zero or positive (logistic): Contained a photo or video	1.442 (1.059-1.963)	.02
Positive count (negative binomial): Contained a photo or video	1.344 (1.021-1.770)	.04

[a] CDC: Centers for Disease Control and Prevention.

[b] AMD: advanced molecular detection.

[c] Hurdle models include two model components: a logistic model and a negative binomial model.

Regarding tweets retweeted by @CDC_AMD, we did not observe any significant difference in the retweet count between different contents in Model F (negative binomial regression model with both Media and Content variables). A likelihood ratio test between Model F and Model G (negative binomial regression model with and without Content variables) found no significant difference between the two (χ^2_5=8.2, *P*=.14). We could not reject the null hypothesis that content was not a confounder, and we therefore selected Model G that was simpler and had a lower AIC (Table 4). There was no difference in the retweet count between a tweet sharing photos or images (or their links) and one that did not (Model G: probability ratio=0.825, 95% CI 0.508-1.369, *P*=.44; Table 5).

Table 4. Number of parameters, log-likelihood, and Akaike Information Criterion for the 2 models that we tested for the corpus of retweets created by the @CDC_AMD Twitter profile.

Model	F	G
Model choice	Negative binomial	Negative binomial
Predictors[a]	Media+Content	Media
Number of parameters	9	4
Log-likelihood	−309.15	−313.27
Akaike Information Criterion	634.2917	632.5408

[a] Media: with a photo or a video (or a link to a photo or a video).

Table 5. Probability ratios for a tweet retweeted by @CDC_AMD being retweeted in a negative binomial regression model that includes both the variable for photo or video attachment and the content variable (Model F), and one without the content variable (Model G).

Explanatory variables of each model	Probability ratio (95% CI)	*P* value
Model F (negative binomial model)		
Contained a photo or video	0.703 (0.424-1.184)	.18
Content		
Referred to the CDC[a] AMD[b] website	Reference	-
Publication in content	3.121 (0.584-25.465)	.21
Webinar in content	0.882 (0.160-7.304)	.89
Bioinformatics in content	0.695 (0.170-3.136)	.62
Scientist in content	1.171 (0.269-6.313)	.84
Miscellaneous content	2.527 (0.890-6.022)	.053
Model G (negative binomial model)		
Contained a photo or video	0.825 (0.508-1.369)	.44

[a] CDC: Centers for Disease Control and Prevention.

[b] AMD: advanced molecular detection.

Discussion

Principal Findings

The AMD program uses Twitter to communicate its accomplishments, provide updates on activities, and share scientific data. This study confirms OAMD's experience that original tweets containing images or videos (or their links) created for the @CDC_AMD Twitter profile received more retweets. The number of retweets was similar across content topics posted by @CDC_AMD. Our case study of a Twitter handle specializing in communicating public health applications of AMD provides concrete evidence that informs public health communication in practice.

As of November 2016, 54.9% (444/809) of @CDC_AMD's followers did not belong to any of the prespecified categories. The a priori–defined categories were categories of target audience specified in OAMD's existing communication strategies. Our results indicated that @CDC_AMD has reached an audience beyond its initial target audience. Anecdotal evidence suggests that these miscellaneous followers could potentially be members of the public who are interested in science and in public health, including university students. Further research will investigate who these people are. By identifying followers who retweeted @CDC_AMD's tweets, @CDC_AMD may engage them for help disseminating information across Twitter.

Our study found that 51.9% (354/682) of original tweets from @CDC_AMD directed users to the website of OAMD. It suggests that OAMD relied on their website to communicate scientific information in details, whereas they used Twitter to alert users to updates of their website contents. OAMD also promoted their scientists' research papers; 19.5% (133/682) original @CDC_AMD tweets fell in that category. Our study also found that @CDC_AMD did tweet on topics other than the prespecified content categories, by both tweeting their own tweets and retweeting others' tweets. It showed scientific and event information not captured in the prespecified content categories defined by OAMD, but that were deemed relevant to their @CDC_AMD followers by OAMD at an ad hoc basis. Future research into the tweets in the miscellaneous category would help OAMD detail these tweets and develop new content categories for the playbook of their routine communication strategies.

In the Context of the Literature

Previous research has identified factors that contributed to more retweets. Suh et al found that the presence of hashtags and URLs in a tweet and the number of followers and friends (users whom one follows) of a Twitter user were positively associated with the probability of a tweet being retweeted [24]. Can et al found that the ratio between followers and friends was highly correlated with retweet count [25]. Soboleva et al noted that hashtags and retweet requests were associated with higher retweet rates, whereas URL links and mentions were associated with lower rates of retweeting [26].

In this study, we focused on the effect of attaching images or videos to tweets on their retweet frequency, in particular, tweets tweeted by @CDC_AMD. According to Soboleva and colleagues [26], images can increase the effect of advertisements partly because they can convey meanings not expressed via words. In their literature review, Soboleva et al [26] also highlighted that images can influence consumer persuasion, have the potential to effect attitude change, and increase recall of advertisements' verbal information. Prior studies on the effect of links to visual cues on retweet frequency using Twitter data collected in 2011 had conflicting results [27,28]. Analyzing tweets associated with tourism in European cities, Bruni et al [27] found that tweets with a link to a photo or a video had more retweets than those without, and tweets linking to a photo had more retweets than tweets linking to a video. On the contrary, Malhotra et al [28] could not identify statistically significant effect of embedded links to websites, photos, or videos on retweet likelihood of tweets tweeted by 47 major commercial brands. However, another study that analyzed tweets from 2009 to 2012, drawn from 298 Twitter profiles operated by 100 top brands, found that links to photos or videos increased the likelihood of a tweet being retweeted [29]. A more recent study analyzing 2014 Twitter data posted by 32 major commercial brands found that photos were consistently associated with higher retweet rates across 3 industries under study (automobile, fast-moving consumer goods, and luxury) [26]. Nevertheless, the same study found that the industry of the brand was an effect modifier for the effect of videos on retweet rates (significant increase in retweet probability for luxury brands, but insignificant effect for the other 2 industries) [26]. Similar research on Sina Weibo (a Chinese social media platform similar to Twitter) also found that multimedia Weibo posts received more reposts, and were reposted by users for a longer period of time, than text-only Weibo posts [30]. In a retrospective observational study, the effect of attaching visual cues to a tweet on its retweet frequency was found to vary across cycles of original tweets with hashtags #CDCGrandRounds and #VitalSigns. The probability ratios of retweet frequency of tweets with visual cues as compared to tweets without visual cues ranged from less than 1 to as high as 34, depending on the topic of the specific CDC Grand Round event or Vital Signs publication [31]. In a prospective, case-control crossover study of visual abstracts (graphics that summarize the main message of a scientific paper), a surgical journal found that the retweet frequency of its tweet that carries both the title of the paper and a visual abstract was 8.4 times that of a tweet with the title of the paper only (92 vs 11 retweets, $P<.001$) [32]. Further evidence was provided by Can et al who identified certain features of an image that were positively correlated with retweet count [25]. Consistent with prior findings, this study on @CDC_AMD adds evidence to a growing literature that attaching visual cues to tweets will increase their retweet frequency, and this practice may enhance Twitter users' engagement with health communication messages promoted by medical and public health professionals.

Limitations

Our study is limited to its cross-sectional design and to 1 CDC Twitter profile. The strength of this study is that we manually coded and statistically analyzed the entire corpus of tweets published by @CDC_AMD. Although our study follows the protocol defined a priori, it lacks the ability to classify 55% of the followers and 80% of contents retweeted by @CDC_AMD that did not fall into any of the prespecified categories. Further research on the followers is warranted. We select retweet frequency as our measure of engagement of a tweet, while acknowledging its limitations. We acknowledge that other measures of impact exist, including "reach," defined as the sum of the potential number of individuals exposed to each retweet of a tweet. While "reach" may account for the number of followers who retweeted a tweet, retweeting is a high level of engagement, and retweet frequency provides solid evidence to our research questions. We did not distinguish videos from photos when we manually coded the visual cues. Future research can further investigate the difference in effect between these 2 types of visual cues for a public health Twitter handle, as compared with results for Twitter handles of major commercial brands [26]. We also did not analyze specific visual features of the photos or videos (as in [25]) that were beyond the scope of our study. There were other potential confounders or effect modifiers [26] that were not included in our study. Given the nature of this study as an observational study, we cannot rule out the possibility of residual confounding by other factors.

Conclusions

As part of its communication strategies, OAMD includes its Twitter handle, @CDC_AMD, as one of its communication channels to its audience. This study underscored the importance of including visual information to build engagements with @CDC_AMD tweets. Following on this brief study conducted in mid-November of 2016, OAMD increased its use of visual cues for tweets, including photos and graphs, and we have seen steady engagement rates during the subsequent 12 months.

On the basis of anecdotal evidence that Twitter is used highly in the biotechnology and biomedical industries, OAMD decided to use this mechanism to target these audiences. On the basis of monthly reviews of Twitter Analytics, audience areas of interest demonstrate that @CDC_AMD hits the intended audiences (eg, 75% are interested in science news, 70% in tech news, and 70% in biotech and biomedical). Further investigation is needed to identify the 55% of @CDC_AMD's followers who did not belong to any of the prespecified categories in this study. These 444 followers could include personal Twitter handles of followers within the target audience and used outside their professional realm. However, a more in-depth study or survey is necessary to determine their interest in @CDC_AMD.

Acknowledgments

The authors thank Dr Martin I Meltzer (Health Economics and Modeling Unit [HEMU], Division of Preparedness and Emerging Infections [DPEI], NCEZID, CDC) for his support in this project. This research project was conducted as part of ICHF's service to HEMU, DPEI, NCEZID, CDC, via an Intergovernmental Personnel Act agreement (#16IPA1609578). ICHF thanks Prof King-Wa Fu (University of Hong Kong), Prof Jingjing Yin (Georgia Southern University), and Dr George Kawai Leung (Emory University) for statistical consultations. ICHF received salary support from HEMU, DPEI, NCEZID, CDC (#16IPA1609578).

Authors' Contributions

ICHF conceived the research idea and drafted the research protocol. AMJ provided critical inputs to the research protocol. NS and ZTHT retrieved the Twitter data. AMJ, LAM, EBB, and MEG manually coded the Twitter contents and followers under the instruction of ICHF. AMJ and LAM completed the descriptive statistical analysis, the inter-rater reliability calculation, and the Poisson regression model. ICHF did the model selection and applied the negative regression models and the hurdle models to the data. AMJ and LAM wrote the first draft of the manuscript. ICHF rewrote the manuscript. ZTHT and AJG critically reviewed and edited the paper. AJG created the feature image. ICHF and ZTHT serve as the guarantors of the paper.

Conflicts of Interest

None declared.

References

1. Fung IC, Tse ZT, Fu KW. The use of social media in public health surveillance. Western Pac Surveill Response J 2015;6(2):3-6 [FREE Full text] [doi: 10.5365/WPSAR.2015.6.1.019] [Medline: 26306208]
2. Taggart T, Grewe ME, Conserve DF, Gliwa C, Roman IM. Social media and HIV: a systematic review of uses of social media in HIV communication. J Med Internet Res 2015 Nov;17(11):e248 [FREE Full text] [doi: 10.2196/jmir.4387] [Medline: 26525289]
3. Overbey KN, Jaykus LA, Chapman BJ. A systematic review of the use of social media for food safety risk communication. J Food Prot 2017 Sep;80(9):1537-1549. [doi: 10.4315/0362-028X.JFP-16-345] [Medline: 28805456]
4. Finch KC, Snook KR, Duke CH, Fu KW, Tse ZT, Adhikari A, et al. Public health implications of social media use during natural disasters, environmental disasters, and other environmental concerns. Nat Hazards 2016 Apr 19;83(1):729-760. [doi: 10.1007/s11069-016-2327-8]
5. Sinnenberg L, Buttenheim AM, Padrez K, Mancheno C, Ungar L, Merchant RM. Twitter as a tool for health research: a systematic review. Am J Public Health 2017 Jan;107(1):e1-e8. [doi: 10.2105/AJPH.2016.303512] [Medline: 27854532]
6. Fung IC, Duke CH, Finch KC, Snook KR, Tseng P, Hernandez AC, et al. Ebola virus disease and social media: a systematic review. Am J Infect Control 2016 Dec 01;44(12):1660-1671. [doi: 10.1016/j.ajic.2016.05.011] [Medline: 27425009]
7. Fung IC, Fu KW, Chan CH, Chan BS, Cheung CN, Abraham T, et al. Social media's initial reaction to information and misinformation on Ebola, August 2014: facts and rumors. Public Health Rep 2016;131(3):461-473 [FREE Full text] [doi: 10.1177/003335491613100312] [Medline: 27252566]
8. Radzikowski J, Stefanidis A, Jacobsen KH, Croitoru A, Crooks A, Delamater PL. The measles vaccination narrative in Twitter: a quantitative analysis. JMIR Public Health Surveill 2016;2(1):e1 [FREE Full text] [doi: 10.2196/publichealth.5059] [Medline: 27227144]
9. Fung ICH, Zeng J, Chan CH, Liang H, Yin J, Liu Z, et al. Twitter and Middle East respiratory syndrome, South Korea, 2015: a multi-lingual study. Infect Dis Health 2018 Mar;23(1):10-16. [doi: 10.1016/j.idh.2017.08.005]
10. Fu KW, Liang H, Saroha N, Tse ZT, Ip P, Fung IC. How people react to Zika virus outbreaks on Twitter? A computational content analysis. Am J Infect Control 2016 Dec 01;44(12):1700-1702. [doi: 10.1016/j.ajic.2016.04.253] [Medline: 27566874]
11. Stefanidis A, Vraga E, Lamprianidis G, Radzikowski J, Delamater PL, Jacobsen KH, et al. Zika in Twitter: temporal variations of locations, actors, and concepts. JMIR Public Health Surveill 2017 Apr 20;3(2):e22 [FREE Full text] [doi: 10.2196/publichealth.6925] [Medline: 28428164]
12. US Centers for Disease Control and Prevention. CDC. 2014 Jan 14. Twitter Guidelines and Best Practices URL: https://www.cdc.gov/socialmedia/tools/guidelines/twitter.html [accessed 2017-12-26] [WebCite Cache ID 6w0OA8z0b]
13. US Centers for Disease Control and Prevention. CDC. Social Media at CDC: Tools: Twitter URL: https://www.cdc.gov/socialmedia/tools/Twitter.html [accessed 2018-02-22] [WebCite Cache ID 6xQhNv92L]
14. US Centers for Disease Control and Prevention. CDC. 2011 Dec 01. Social Media Guidelines and Best Practices URL: https://www.cdc.gov/socialmedia/tools/guidelines/pdf/microblogging.pdf [accessed 2018-02-22] [WebCite Cache ID 6xQhWW3VW]
15. Bedrosian SR, Young CE, Smith LA, Cox JD, Manning C, Pechta L, et al. Lessons of risk communication and health promotion - West Africa and United States. MMWR Suppl 2016 Jul 08;65(3):68-74. [doi: 10.15585/mmwr.su6503a10] [Medline: 27386834]

16. Lazard AJ, Scheinfeld E, Bernhardt JM, Wilcox GB, Suran M. Detecting themes of public concern: a text mining analysis of the Centers for Disease Control and Prevention's Ebola live Twitter chat. Am J Infect Control 2015 Oct 01;43(10):1109-1111. [doi: 10.1016/j.ajic.2015.05.025] [Medline: 26138998]
17. Glowacki EM, Lazard AJ, Wilcox GB, Mackert M, Bernhardt JM. Identifying the public's concerns and the Centers for Disease Control and Prevention's reactions during a health crisis: an analysis of a Zika live Twitter chat. Am J Infect Control 2016 Dec 01;44(12):1709-1711. [doi: 10.1016/j.ajic.2016.05.025] [Medline: 27544795]
18. US Centers for Disease Control and Prevention. Twitter. 2018. @CDCgov URL: https://twitter.com/CDCgov [accessed 2018-02-22] [WebCite Cache ID 6xQl2QMLC]
19. US Centers for Disease Control and Prevention. CDC. 2017 Aug 12. AMD: Advanced Molecular Detection URL: https://www.cdc.gov/amd/pdf/amd-at-a-glance-final-508.pdf [accessed 2017-08-12] [WebCite Cache ID 6sfahjwnY]
20. Guinn AJ. A Tweet, a Blog and an Unusual Bug: how social media and open source data sharing led to an international collaboration. 2016 Presented at: Poster presentation at the National Conference on Health Communication, Marketing & Media; 2016; Atlanta, GA.
21. US Centers for Disease Control and Prevention. CDC. 2017 Aug 12. Advanced Molecular Detection URL: http://www.cdc.gov/amd/ [accessed 2017-08-12] [WebCite Cache ID 6sfb3zSar]
22. Viera AJ, Garrett JM. Understanding interobserver agreement: the kappa statistic. Fam Med 2005 May;37(5):360-363 [FREE Full text] [Medline: 15883903]
23. Zeileis A, Kleiber C, Jackman S. Regression models for count data in R. J Stat Softw 2008;27(8):1-25 [FREE Full text]
24. Suh B, Hong L, Pirolli P, Chi EH. Want to be retweeted? Large scale analytics on factors impacting retweet in Twitter network. 2010 Aug 20 Presented at: IEEE Second International Conference on Social computing (socialcom); 2010 Aug 20; Minneapolis, MN p. 177-184 URL: https://doi.org/10.1109/SocialCom.2010.33
25. Can EF, Oktay H, Manmatha R. Predicting retweet count using visual cues. 2013 Oct 27 Presented at: The 22nd ACM International Conference on Information & Knowledge Management; 2013 Oct 27; San Francisco, CA, USA p. 1481-1484. [doi: 10.1145/2505515.2507824]
26. Soboleva A, Burton S, Mallik G, Khan A. 'Retweet for a Chance to...': an analysis of what triggers consumers to engage in seeded eWOM on Twitter. J Mark Manage 2017 Sep 13;33(13-14):1120-1148. [doi: 10.1080/0267257X.2017.1369142]
27. Bruni L, Francalanci C, Giacomazzi P. The role of multimedia content in determining the virality of social media information. Information (Basel) 2012 Jul 25;3(4):278-289. [doi: 10.3390/info3030278]
28. Malhotra A, Malhotra C, See A. Sloanreview.mit.ed. How to get your messages retweeted URL: https://sloanreview.mit.edu/article/how-to-get-your-messages-retweeted/ [accessed 2018-02-22] [WebCite Cache ID 6xQjFYI2a]
29. Araujo T, Neijens P, Vliegenthart R. What motivates consumers to re-tweet brand content? The impact of information, emotion, and traceability on pass-along behavior. J Advert Res 2015;55(3):284-295. [doi: 10.2501/JAR-2015-009]
30. Zhao X, Zhu F, Qian W, Zhou A. Impact of multimedia in Sina Weibo: popularity and life span. In: Li J, Qi G, Zhao D, Nejdl W, Zheng HT, editors. Semantic Web and Web Science. New York, NY: Springer; 2013:55-65.
31. Jackson AM, Mullican LA, Yin J, Tse ZT, Liang H, Fu KW, et al. #CDCGrandRounds and #VitalSigns: a Twitter analysis. Ann Glob Health 2018 (forthcoming).
32. Ibrahim AM, Lillemoe KD, Klingensmith ME, Dimick JB. Visual abstracts to disseminate research on social media: a prospective, case-control crossover study. Ann Surg 2017 Dec;266(6):e46-e48. [doi: 10.1097/SLA.0000000000002277] [Medline: 28448382]

Abbreviations

AIC: Akaike Information Criterion
AMD: advanced molecular detection
APHL: Association of Public Health Laboratories
CDC: Centers for Disease Control and Prevention
DPEI: Division of Preparedness and Emerging Infections
HEMU: Health Economics and Modeling Unit
NCEZID: National Center for Emerging and Zoonotic Infectious Diseases
OAMD: Office of Advanced Molecular Detection
RQ: research question
UTC: Universal Coordinated Time

Edited by T Sanchez; submitted 12.08.17; peer-reviewed by K Abbas, S Li; comments to author 06.10.17; revised version received 30.12.17; accepted 12.01.18; published 02.04.18.

Please cite as:
Fung ICH, Jackson AM, Mullican LA, Blankenship EB, Goff ME, Guinn AJ, Saroha N, Tse ZTH
Contents, Followers, and Retweets of the Centers for Disease Control and Prevention’s Office of Advanced Molecular Detection (@CDC_AMD) Twitter Profile: Cross-Sectional Study
JMIR Public Health Surveill 2018;4(2):e33
URL: http://publichealth.jmir.org/2018/2/e33/
doi:10.2196/publichealth.8737
PMID:29610112

Original Paper

Using Social Media to Generate and Collect Primary Data: The #ShowsWorkplaceCompassion Twitter Research Campaign

Wendy Clyne[1], PhD; Sally Pezaro[2], PhD; Karen Deeny[1], PhD; Rosie Kneafsey[2], PhD

[1]Hope for the Community, Community Interest Company, Coventry, United Kingdom

[2]School of Nursing, Midwifery and Health, Coventry University, Coventry, United Kingdom

Corresponding Author:
Wendy Clyne, PhD
Hope for the Community, Community Interest Company
The Enterprise Hub
Whitefriars Street
Coventry, CV1 2DS
United Kingdom
Phone: 44 24 7765 9542
Email: wendy.clyne@h4c.org.uk

Abstract

Background: Compassion is a core value embedded in the concept of quality in healthcare. The need for compassion toward healthcare staff in the workplace, for their own health and well-being and also to enable staff to deliver compassionate care for patients, is increasingly understood. However, we do not currently know how healthcare staff understand and characterize compassion toward themselves as opposed to patients.

Objective: The aim of this study was to use social media for the generation and collection of primary data to gain understanding of the concept of workplace compassion.

Methods: Tweets that contained the hashtag *#ShowsWorkplaceCompassion* were collected from Twitter and analyzed. The study took place between April 21 and May 21, 2016. Participants were self-selecting users of the social media service Twitter. The study was promoted by a number of routes: the National Health Service (NHS) England website, the personal Twitter accounts of the research team, internal NHS England communications, and via social media sharing. Participants were asked to contribute their views about what activities, actions, policies, philosophies or approaches demonstrate workplace compassion in healthcare using the hashtag #ShowsWorkplaceCompassion. All tweets including the research hashtag #ShowsWorkplaceCompassion were extracted from Twitter and studied using content analysis. Data concerning the frequency, nature, origin, and location of Web-based engagement with the research campaign were collected using Bitly (Bitly, Inc, USA) and Symplur (Symplur LLC, USA) software.

Results: A total of 260 tweets were analyzed. Of the 251 statements within the tweets that were coded, 37.8% (95/251) of the statements concerned Leadership and Management aspects of workplace compassion, 29.5% (74/251) were grouped under the theme related to Values and Culture, 17.5% (44/251) of the statements related to Personalized Policies and Procedures that support workplace compassion, and 15.2% (38/251) of the statements concerned Activities and Actions that show workplace compassion. Content analysis showed that small acts of kindness, an embedded organizational culture of caring for one another, and recognition of the emotional and physical impact of healthcare work were the most frequently mentioned characteristics of workplace compassion in healthcare.

Conclusions: This study presents a new and innovative research approach using Twitter. Although previous research has analyzed the nature and pattern of tweets retrospectively, this study used Twitter to both recruit participants and collect primary data.

(JMIR Public Health Surveill 2018;4(2):e41) doi:10.2196/publichealth.7686

KEYWORDS

work engagement; health personnel; empathy; attitude of health personnel

Introduction

Using Twitter for Health-Related Research

An increasing number of studies have used Twitter for health-related research. Twitter has been used by researchers not just as a mechanism for dissemination of research findings [1,2] but also as a source of data that is available in the public domain [3]. Research has been particularly fruitful in the areas of public health and infectious diseases where Twitter has been used, for example, to track geographically reports of illness related to consumption of restaurant foods [4] and to explore whether Twitter can be an effective tool for the surveillance of communicable diseases spread at large festival and public gatherings [5]. Both the pattern and the content of tweets about influenza during the H1N1 pandemic were examined to explore how Twitter is used to share health information [6]. The authors highlight the potential for Twitter-based research to aid our understanding of the predictors of uptake of health interventions, such as vaccination, following public health campaigns. The content of tweets was also examined in a study concerned with tweet content in the day-to-day Twitter communications of health professionals [7].

Researchers have also used online social networking sites as a mechanism for recruiting participants to traditional research studies. Online recruitment has been used for recruiting underserved and hard-to-reach groups that may be difficult to engage by more traditional means [8]. Research studies using this approach have conducted online research recruitment campaigns across a number of online platforms simultaneously (Facebook, Twitter, LinkedIn) [8] or across a single platform (Twitter) [9,10] to recruit participants to take part in research using traditional data collection methods. For the Twitter component of their multiplatform recruitment strategy, Yuan et al [8] created a research account and associated hashtags to recruit individuals to an online survey about HIV clinical outcomes. They found social media to be indispensable for recruitment and were particularly taken with the efficiency and cost-effectiveness of their approach.

Crowdsourcing Primary Data

In our study, we combined the benefits of Twitter as a participant recruitment method and also used Twitter as a means for generating primary data. To the best of our knowledge, this is the first study to both recruit participants and collect primary data using Twitter.

Our study concerned compassion in healthcare. Compassion for patients is a central tenet of the delivery of care and treatment to patients to prevent and treat ill-health. Compassion has been defined as “the combination of underpinning emotions (such as sympathy and empathy), with altruistic values, (particularly a desire to help others), which together motivate an individual to take action, which would ultimately be experienced as ‘care’ by the recipient” [11]. Health staff who show compassion toward patients tend to be more effective at delivering care [12]. Latterly, the concept of compassion has been applied to the way in which healthcare is designed and delivered at the organizational level [13]. Thus, the recipient of compassion in healthcare may be staff, not just patients. Given that healthcare employers have a responsibility to ensure the health and safety of their staff, and the evidence that positive staff experience is associated with positive patient experience [14], compassion toward staff potentially benefits both staff (directly) and patients (indirectly). Although the concept of compassion toward patients has been explored and defined, we do not, as yet, understand the nature and characteristics of compassion as it applies to healthcare staff. In our study, which was conducted using Twitter, we invited healthcare staff to tell us what #ShowsWorkplaceCompassion. Twitter users working in any role, setting, or organization in healthcare across England were invited to participate in a research campaign to explore the characteristics of workplace compassion as they perceived it.

Methods

Design

This qualitative study reports the analysis of tweets posted by users of the online social media platform Twitter and uses template analysis to explore what Twitter users perceive shows workplace compassion in a healthcare setting. Statistical data were also captured about the frequency, nature, and origin of online engagements with the study.

Participant Recruitment

A study webpage was uploaded to the National Health Service (NHS) England website. This invited people to participate in the research study as follows: “We want to know what you think a ‘compassionate healthcare organisation’ looks like. It’s easy to take part. Just tweet your views about what activities, actions, policies, philosophies, approaches, that demonstrate that a healthcare organisation is compassionate using the hashtag #ShowsWorkplaceCompassion. Your contributions—your tweets—might be based on your own experiences of working in healthcare, your experience of providing or commissioning healthcare or based on your general views about what healthcare organisations are about. There are no right or wrong ideas or perspectives, it is your ‘vision’ we are interested in.” Potential participants were advised that they could tweet as many times as they chose and were given a reminder to always use the hashtag #ShowsWorkplaceCompassion. A link was provided to a participant information leaflet. This was shared throughout the data collection period by the research account (@NHSStaffExp) and the research team. The research study was promoted on social media from the individual Twitter accounts of the research team and the NHS England Twitter account. Key influencers on Twitter were contacted ahead of study launch and invited to promote the study. Key influencers included prominent online health and social care individuals and organizations such as the “WeCommunities,” patient groups, research centers, NHS organizations, healthcare commentators, commissioners, and healthcare leaders. Key influencers were told about the purpose of the study and how we envisaged the hashtag would be used by participants:

> *In a few weeks time, we will be launching a twitter account and the hashtag #ShowsWorkplaceCompassion. We will be inviting people to tweet what they think shows workplace compassion for staff by tweeting sentences that end*

> *with #ShowsWorkplaceCompassion (e.g., 'letting me leave early when my son was ill #ShowsWorkplaceCompassion'; 'senior management talking to staff #ShowsWorkplaceCompassion', or 'I know my manager cares about me and that #ShowsWorkplaceCompassion'. We do hope you will join us in finding out and tweeting about what #ShowsWorkplaceCompassion.*

Throughout the data collection period, the research account tweeted daily to encourage Twitter users to participate in the study. These tweets targeted Twitter users working in healthcare. Promotional tweets were also used to recruit participants. Each promotional tweet was seen by between approximately 1100 and 4300 Twitter users. Links to the study webpage were also shared via other social media platforms.

Procedure

Before data collection, the research hashtag "#ShowsWorkplaceCompassion" was registered with Symplur (Symplur LLC, USA) as a healthcare hashtag to monitor the frequency and number of tweets shared and seen by Twitter users. The URL links to both the study information webpage and the participant information webpage were also logged via Bitly (Bitly, Inc, USA) software to record the geographical location and number of online engagements with both webpages.

The Twitter research account was monitored between 8 am and 11 pm (GMT) for 7 days a week over the 4-week study period in Spring 2016. The biographical information for this research account was listed as "Research account: (@CovUni_CTEHR & @NHSEngland). Let us know what #ShowsWorkplaceCompassion. tweets used for research. See full details→URL Link." URL links to both the study information webpage and the participant information leaflet webpage were visible at all times on the research account's profile page.

Hootsuite software was used to schedule and send 9 tweets from the research account between 6 am and 10 pm each day over the period of data collection. Tweets sent by the research account such as "RU a healthcare worker? Have you just finished a nightshift? Did you experience anything that #ShowsWorkplaceCompassion? Let us know! Pls RT" were designed to be promotional and engaging in nature. Links to both the study information webpage and the participant information webpage were also shared.

The permanent "pinned" tweet that remained pinned to the top of the research account's Twitter page throughout the data collection period was "Our #ShowsWorkplaceCompassion research explores staff 'visions' of compassionate healthcare workplaces. See here → URL Link." The research account also published other tweets on an impromptu basis to answer specific questions and to offer guidance as necessary in response to questions.

Twitter accounts related to healthcare were "followed" by the research account. This account also "liked" and "retweeted" (shared a copy of) every tweet that contained the hashtag "#ShowsWorkplaceCompassion." At the end of the data collection period, the pinned tweet on the research account was changed to "Data collection 4 #ShowsWorkplaceCompassion project has finished→ URL Link Thx 2 all participants. Results 2B published here."

Data Collection

Qualitative data were collected via tweets posted on Twitter. Statistical data about the nature, frequency, number, and geographical location of research related to online engagements were downloaded using Symplur and Bitly software. All tweets posted during the 4-week data collection period containing the hashtag #ShowsWorkplaceCompassion were downloaded directly from the Twitter platform into Microsoft Excel.

Data Analysis

Symplur software was used to identify the frequency and number of tweets seen and shared by Twitter users. The geographical location and number of known online engagements with both the study information webpage and the participant information webpage were identified using Bitly software.

The downloaded tweets were initially sorted into chronological order to allow the identification and removal of any duplicated material. Any tweets originating from either the research team or the research account were also excluded from the dataset. Additionally, tweets that simply promoted this research or referred the research team to secondary data sources were also removed at this point.

A total of 110 tweets (12.9%) that included the research hashtag expressed support for the research project and encouraged other followers to share their ideas. Although these were considered to contribute to the success of this research, they were not suitable for analysis and were therefore excluded from the analysis. Of the remaining tweets, 8.8% (n=22) either partially or completely referred to patients as the target of compassion rather than staff. These references were therefore excluded from the analysis. While 1.8% (n=16) of tweets containing the research hashtag directed the research team toward secondary sources, 1.4% (n=12) of tweets were retweets (shared copies) of other tweets. These tweets were also excluded from the analysis.

Subsequently, content analysis was conducted for each tweet. An adapted template analysis approach was used [15]. The coding scheme from a focus group about workplace compassion with healthcare commissioners (reported elsewhere) was used as the basis for analysis of data in this study. Four themes—(1) feelings, (2) actions, (3) circumstances, and (4) organizational factors—served as the template for analysis. As the analysis progressed, the template was adapted to the emerging data with theme names changed to "Values and Culture," "Activities and Actions," "Personalized Policies and Procedures," and "Leadership and Management." All content analysis was conducted by 1 coauthor (SP) and then checked and confirmed by a second coauthor (WC). Areas of disagreement between reviewers were resolved by consensus on a case-by-case basis. Reviewers were blind to identifying information about participants during content analysis.

Statements that described feelings such as having a common purpose or feeling valued as incidences of workplace compassion were coded under the appropriate subtheme within the first theme (Values and Culture). Any actions described as incidences of workplace compassion within statements were coded under the appropriate subtheme within the second theme (Activities and Actions). Individual circumstances covered by policies and procedures that were perceived to show workplace compassion were coded under the appropriate subtheme within the third theme (Personalized Policies and Procedures). Statements that described the wider organizational, practical, and/or leadership principles as instances of workplace compassion were coded under the appropriate subtheme within the final theme (Leadership and Management).

Results

Recruitment and Social Media Activity

Following a 4-week data collection period, 850 tweets that included the research hashtag were collected from Twitter. The study webpage was accessed 177 times, and the participant information webpage was accessed 406 times. The participant information leaflet was accessed more frequently than the general study information webpage. Both were accessed more frequently in the first half of the research campaign.

The number of tweets posted using the research hashtag over the 4-week data collection period were plotted from week 1 to week 4 (Figure 1). Tweet activity was the highest during the first week of the research campaign and then dropped during the remaining 3 weeks.

Table 1 summarizes the number of tweets that were included in the content analysis and reasons for exclusion, as well as information about recruitment and social media activity data. The research team noted that there were 2 newly created Twitter accounts that had not tweeted before tweeting with the study hashtag. However, the majority of tweets were posted by established Twitter users. The research hashtag made 6,037,026 online impressions, that is, the research hashtag was delivered to Twitter users' Twitter streams 6,037,026 times. A total of 452 tweets were generated by the research team or the research account to promote the study, resulting in generation of 260 tweets that were appropriate for content analysis.

Content Analysis

Content analysis was used to analyze the 260 tweets concerning what shows workplace compassion in healthcare. This process resulted in the development of 4 themes and 19 subthemes (Table 2).

Of the 251 statements within the tweets that were coded, 37.8% (n=95) of statements concerned Leadership and Management aspects of workplace compassion, 29.5% (n=74) were grouped under the theme related to Values and Culture, 17.5% (n=44) of statements related to Personalized Policies and Procedures that support workplace compassion, and 15.2% (n=38) of statements concerned Activities and Actions that show workplace compassion.

Leadership and Management

Approximately one quarter of statements (n=26) within this theme referred to an embedded organizational culture of caring for one another as a significant feature of workplace compassion. One tweet reflects this description of workplace compassion as "Taking time to understand everyone's issues and then pulling together."

An organizational culture of speaking openly to learn from mistakes comprised a subtheme within the Leadership and Management theme. This openness was referred to from both an organizational and individual perspective. On an individual basis, supporting staff to "speak out safely" was tweeted as a demonstration of workplace compassion. In relation to organizational responses to adverse events, organizational learning and openness were considered to show workplace compassion. One tweet captured this view as follows: "Learning from mistakes instead of punishing #ShowsWorkplaceCompassion." Another tweet captured how "Regulation that engages with the reality not the fantasy #ShowsWorkplaceCompassion."

Figure 1. Tweet activity for the #ShowsWorkplaceCompassion hashtag.

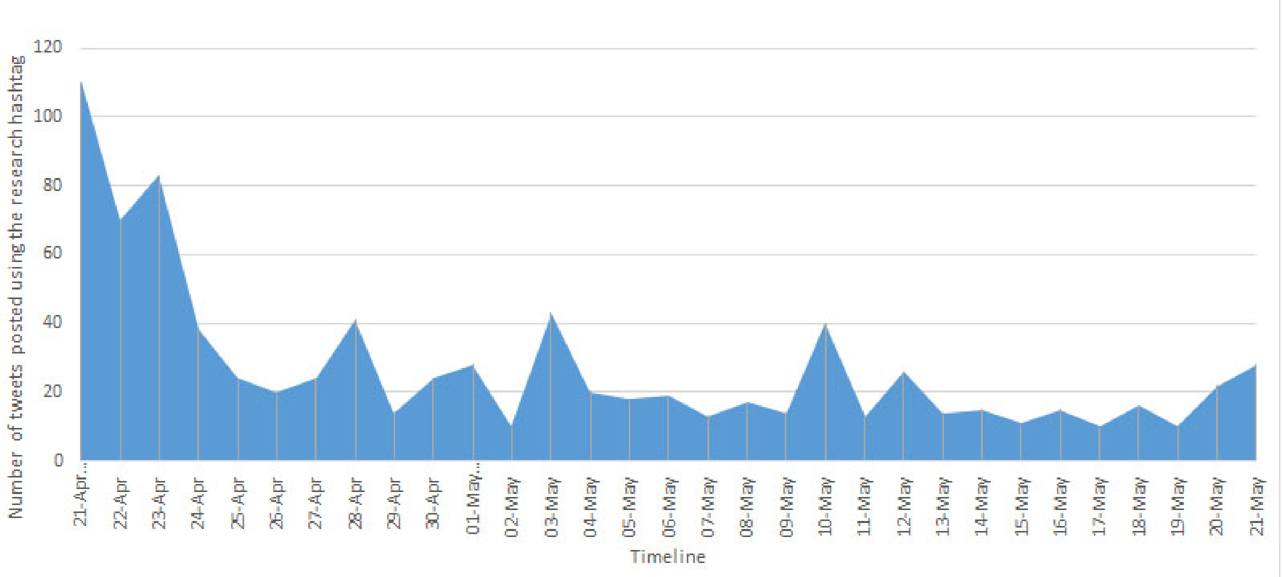


Table 1. Summary of included and excluded tweets, recruitment, and Twitter activity.

Included and excluded tweets	Number
Total tweets extracted	850
Tweets originating from the research account removed	295
Tweets originating from the research team removed	157
Tweets promoting the research removed	110
Tweets leading toward secondary sources removed	16
Retweets removed	12
Total tweets for analysis	260
Research study information webpage visits	177
Research study information webpage visitor locations	
United Kingdom	106
Germany	43
United States	20
Unknown	8
Referring channels to research study information webpage	
Unidentifiable	79
Twitter	66
Links	25
Unknown	7
Participant information webpage visits	406
Participant information webpage visitor locations	
United Kingdom	247
Germany	116
United States	14
Unknown	29
Referring channels to participant information webpage	
Unidentifiable	149
Facebook	109
Links	91
Unknown	57
Tweet activity for the #ShowsWorkplaceCompassion hashtag	
Number of hashtag impressions	6,037,026
Number of Twitter accounts using the hashtag	645

No blame and no bullying leadership and management were also referred to as characteristics of workplace compassion. Action taken against bullying was seen as characteristic of workplace compassion, as tweets stated how “meaningful action & education to identify & act on bullying #ShowsWorkplaceCompassion” and also the “elimination of bullying and bullies…” The identification of “No belittling of staff” and “No gossiping and no bullying” were also put forward as evidence of workplace compassion. Instead, there was a call to “treat others as we would wish to be treated.”

One statement stating that “leaders must role model positive behaviours and values” is representative of many other statements within this theme, which describes how “Compassion is a key component of leadership” and that “Leadership and the right culture is essential” in the demonstration of workplace compassion. Comparable tweets described workplace compassion as “Starting from the top.” Here, compassionate leadership was seen as “fundamental,” yet it was also recognized that organizationally, all staff “living the values of compassion” showed workplace compassion.

Table 2. Themes and subthemes for what #ShowsWorkplaceCompassion.

Theme and subtheme, n (%)	Count
Leadership and Management, 95 (37.8)	
Embedded organizational culture of caring for one another	26
Speaking openly to learn from mistakes	19
No blame/no bullying management	18
Inspiring leaders and collective leadership	14
Financial investment in staff	10
Recognize humanity and diversity	8
Values and Culture, 74 (29.5)	
Common purpose in a team	19
Feeling valued	18
Being heard	16
Enjoying work	9
Engaged	8
Use of caring language	4
Personalized Policies and Procedures, 44 (17.5)	
Recognition of the emotional and physical impact of healthcare work	22
Recognition of nonwork personal context	10
Work/life balance is respected	6
Respecting the right to breaks	4
Being treated well when unwell	2
Activities and Actions, 38 (15.2)	
Small gestures of kindness	30
Provision of emotional support	8

Values and Culture

The first subtheme comprising the highest number of statements about Values and Culture concerned sharing a sense of common purpose within a team. “Compassionate teams are key. Do we care about, value and respect everyone in the team?” is typical of tweets within this theme, along with “clear values shared by all” and “proud to be part of a team.”

This subtheme was closely followed in frequency by one in which workplace compassion is demonstrated by feeling trusted, valued, appreciated, and respected. Tweets such as “Trusting me to do the right thing #ShowsWorkplaceCompassion” and “Nurturing good people by valuing, respecting, rewarding them #ShowsWorkplaceCompassion.” One tweet also described feeling heard as well as trusted, as “Being trusted and being listened to is important #ShowsWorkplaceCompassion.” Being listened to was reported to show workplace compassion and was expressed in other tweets such as “If [staff] feel heard & empowered they’ll treat others [the] same.”

Following on from this, feeling engaged and able to have influence in the workplace were described as demonstrating workplace compassion. This is illustrated in the following tweet: “staff can have a say in strategy and what staff say can actually change the way the NHS is managed.” Enjoying the workplace was a characteristic of workplace compassion for some participants. Examples of this include “Having fun at work important—#ShowsWorkplaceCompassion” and “laughter... #ShowsWorkplaceCompassion.”

Use of sensitive and caring language was referred to as demonstrating workplace compassion in tweets such as “Remember words can be weapons. A team that avoids belittling language #ShowsWorkplaceCompassion” and “Recognising the importance of language #ShowsWorkplaceCompassion…‘we’ ‘us’ ‘colleagues’ ‘people’ not ‘you’ ‘they’ ‘staff’ ‘human resources’.”

Personalized Policies and Procedures

This theme concerned the compassionate application of human resources policies and procedures on an individual basis and when appropriate. When circumstances are challenging for staff, the compassionate operation of policies and procedures to adapt to this were given as examples of workplace compassion. Flexible working patterns that support work-life balance, caring for other relatives, family life, and responses to family crises were also cited as a demonstration of workplace compassion. In this context, tweets described how showing an interest in the home lives and circumstances of staff showed workplace compassion: “my team were fab after the death of my husband”

and "on the loss of his mum colleagues bought my husband a tree to remember her."

The recognition of the emotional and physical impact of caring within an organization was also firmly established as characteristic of workplace compassion, with 1 tweet stating that "We need to talk about compassion fatigue and burnout." Moreover, this recognition was often described as providing time and space to talk and listen. One statement reflects upon this as "space and time for staff to listen and talk about difficult emotions."

Some tweets described practical ways in which emotional needs may be met. These referred to mindfulness training, chaplaincy, restorative resilience, and the provision of practitioner health programs and Schwartz rounds.

In reference to staff experiencing episodes of ill health, 1 tweet stated that "Supporting colleagues through mental health crisis and recognising that adverse behaviour is not misconduct" showed workplace compassion. Supporting colleagues through a phased return after sickness absence was proposed in another tweet as showing workplace compassion in challenging circumstances. Overall, workplace compassion is described in this subtheme of tweets as generally "recognising what matters to one another," with a majority of tweets referring to an interpersonal recognition of each other's personal circumstances.

Activities and Actions

A total of 30 statements in one subtheme concerned general pleasantries or small acts of kindness. Examples of these were given in tweets such as "asking how are you today especially to junior staff and a smile goes a long way (#ShowsWorkplaceCompassion)" and "A moment to make a cup of tea, to share a difficult time #ShowsWorkplaceCompassion." Other tweets referred to other small acts of kindness, such as "a note telling me to stay strong and some flowers" and described acts such as this as "making a big difference."

Discussion

Principal Findings

The 260 tweets analyzed in this study describe the characteristics of workplace compassion for healthcare staff, according to Twitter users participating in this study. Workplace compassion can be seen in Values and Culture, Activities and Actions, Personalized Policies and Procedures, and Leadership and Management. Content analysis showed that small acts of kindness, an embedded organizational culture of caring for one another, and recognition of the emotional and physical impact of healthcare work were the most frequently mentioned characteristics of workplace compassion in healthcare.

Implications for Healthcare Staff and Organizations

Twitter users were able to describe and convey the characteristics of workplace compassion for healthcare staff within the confines of the 140-character limit of Twitter. This study provides evidence about the characteristics of compassion as it applies to healthcare staff as recipients of compassionate behavior. Future research could usefully explore the extent to which healthcare staff perceive themselves to experience workplace compassion, as characterized here, and whether this varies across types of organizations and staff roles. Given the interplay between staff and patient experience, the experience of workplace compassion may be particularly key for frontline staff [16,17].

The Values and Culture that reflect workplace compassion were clearly articulated by participants in this study: a sense of common purpose with others, having a voice and being heard, and feeling valued in the workplace.

The research presented here potentially informs actions by healthcare organizations to improve the experience of staff and the ways in which healthcare organizations measure staff experience. Compassionate leadership, defined as leadership that is adaptive, shared, and distributed [13], has been the focus of research concerned with healthcare staff as recipients of compassion. Participants in this study confirm the importance of leadership and management in achieving compassion in the workplace, with leadership and management comprising the largest category of tweets. As a consequence of this study, healthcare leaders and managers know that compassionate leadership can be achieved by supporting staff to care for one another, taking action against bullying, creating a learning culture, and inspiring others as a role model of compassion in the workplace.

For many participants in this study, workplace compassion is embodied in the interpersonal activities and actions of colleagues, from taking the time to make a drink for a colleague to appreciation of the wider context of a staff member's personal circumstances. In this regard, compassion toward staff, similar to compassion toward patients, has a significant interpersonal behavior component [18]. When staff surveys (such as the NHS staff survey) do cover interpersonal actions and behaviors, these tend to focus on the actions of immediate managers and senior managers rather than peers and colleagues in the workplace. In so doing, such surveys fail to capture a significant interpersonal element of staff experience. Although leadership is clearly an important element of workplace compassion, this study indicates that staff surveys should be adapted to capture the activities and actions that take place between colleagues to fully capture the range of actors involved in achieving workplace compassion and positive staff experience.

Although many health and care staff gain satisfaction from the provision of compassionate care, it can also be emotionally, physically, and mentally demanding for staff. Errors in health and care also take a toll on health and care staff in addition to the impact that errors have on patients and their families. The term "second victim" is sometimes used to describe the impact on the workforce when harm occurs to patients in the receipt of healthcare services. A characteristic of workplace compassion identified in this study is the application of policies and procedures to individuals in ways that acknowledge this reality for staff.

For participants in our study, workplace compassion exists in the details of policies and procedures but is enacted interpersonally between staff, and not solely in interactions with staff in hierarchical positions of authority and power. Initiatives that support healthcare staff to be compassionate toward and between their colleagues in the ways described by participants in this study, including at times of work pressure and when staff have experienced an emotional and physical impact of their work, could be a more effective way to increase the experience of workplace compassion than the development of workplace policies and statements extolling the virtues of compassion in the workplace. As we know that organizations that prioritize staff health and well-being perform better, with improved patient satisfaction, stronger quality scores, better outcomes, higher levels of staff retention, and lower rates of sickness absence [19], and that healthcare organizations can take action that improves support for staff and staff well-being [20], an understanding of the characteristics of workplace compassion is key to making the organizational and cultural change necessary.

Methodological Issues

Our study, which took place over a 1-month period, saw 645 Twitter accounts participate in the #ShowsWorkplaceCompassion research hashtag, generating 260 tweets that could be analyzed to answer our research question. Previous research has compared the efficiency of Web-based recruitment and data collection with paper-based and postal methods. Ebert et al [21] found that Web-based recruitment and data collection generated more complete data and was more cost-effective than postal recruitment and paper based data collection by a factor of 10. Processing digital questionnaires was half the cost of handling postal questionnaires. A randomized controlled trial of recruitment and data collection methods for a survey of doctors found that Web-based methods were cheaper than combined Web-based and postal methods, with no evidence of increased response bias but lower response rates [22]. This study offers similar cost efficiencies through Web-based data collection. Tweets from the research account were automated and scheduled, and daily monitoring of hashtag use and retweeting of contributions to the study could easily be fitted around other research activities.

Previous studies with Web-based data collection recruited potential participants digitally using existing email distribution lists. In contrast to previous studies, this study enabled open recruitment of participants for whom no contact details were available with the research team. Twitter provides the ability to crowdsource primary data. However, the ease and openness of this method of data collection is countered by a challenge to the generalizability of our findings, given that participants were existing Twitter users willing to respond to the research campaign and thus highly self-selecting. Researchers considering using this method need to consider the trade-off between these factors when selecting their methodological approach.

Twitter offers the additional benefit to participant recruitment of easy reach across a wide geographical area. The study webpage was visited by a sizable minority of international participants, despite active promotion only within England. Future research studies seeking an international dataset may find that using Twitter as a recruitment and data collection medium in itself promulgates the research study. Twitter offers a particularly viable route to participant recruitment and primary data collection for research studies with an international focus.

Participation in the study decreased rapidly after the first week of the data collection period. Promotional activity at the start of the study was successful in driving visits to the research account and generating data. Additional promotional activity in later weeks may have been beneficial for maintaining recruitment and momentum. Scheduled tweets from the research account alone were insufficient to maintain the initial level of data collection.

An alternative option for data collection would have been to set up a Twitter chat for data collection. Akin to a focus group, a Twitter chat is a focused discussion of a particular pre-arranged topic or topics to which Twitter users are invited to contribute. Tweet chats often last about an hour. Chai et al [23] tested this approach for primary data collection and found the use of an existing Twitter chat forum essential for securing participants. However, in our study, we were not interested specifically in interaction between Twitter users. Indeed, we conducted a traditional focus group before this study to assist us to formulate this research study, and thus, a Twitter focus group would have duplicated this effort rather than extend our understanding of the topic. Our approach did not rely upon an existing forum of participants active in an existing Twitter chat, but instead sought participants from a wider, open base across Twitter.

The application of formal ethical approval processes and permissions in research using Twitter is variable. In their systematic review, Sinnenberg et al [3] report that about a third of published research studies refer to ethics board approval. As much as Twitter research uses retrospective analysis of existing information already in the public domain, researchers have not necessarily needed to seek formal ethical approval. Although there may be some ambiguity about the extent to which tweets and other social media communications exists in the public versus private domain, and thus the ethical issues that pertain to their use in research [24], there was no such ambiguity in our study. Twitter users were explicitly invited to tweet to generate primary research data in our study. We therefore did seek and secure institutional ethical approval for our approach to participant recruitment. The nature of Twitter does impose several limitations on researchers that necessitates a creative response to ensure consent. As communication on Twitter was limited to 140 characters per tweet, and so participant information details cannot easily be tweeted, we opened a specific account on Twitter for research purposes, and for this study, to be able to share links to the study participant information sheet, hosted elsewhere, in the Twitter account biography and header. The research account was also able to tweet the link to the study participant information routinely at intervals during the data collection period. As the account was new and not a pre-existing Twitter account repurposed for this study, we can be reasonably sure that the research purpose of the account, the hashtag, and the tweets using the hashtag were clear to potential and actual participants. The frequency with which the participant information webpages were accessed is reassuring that many potential study participants did indeed access this information. However, we cannot determine whether all study participants accessed this information. The creation of a specific Twitter account has the additional benefit that we can easily use the Twitter account to share the results of the study with participants and other people that followed the account. The account currently has just under a thousand followers, and we can tweet the details of this research paper directly to those interested in the research.

Conclusions

Twitter can be an effective means for the recruitment of research participants and the collection of primary data about health-related matters. Our study recruited healthcare staff who use Twitter to describe the characteristics of workplace compassion. We found that small acts of kindness, an embedded organizational culture of caring for one another, recognition of the emotional and physical impact of healthcare work, feeling valued, a sense of common purpose within a team, and being able to speak openly to learn from mistakes characterized compassion toward healthcare staff in the workplace. The need to capture effectively the interpersonal elements of workplace compassion in the measurement of staff experience is discussed, for organizations to seek to achieve positive staff experience in healthcare workplaces.

Acknowledgments

The authors would like to acknowledge the support of Sara Latham, Staff Experience Programme Coordinator, NHS England, in assisting with the conduct of the study. This study was funded by NHS England.

Authors' Contributions

WC was the principal investigator. WC, SP, KD, and RK designed the study. SP was responsible for social media management, and WC and SP for data analysis and data interpretation. All the authors helped to draft the manuscript and approved the final version.

Conflicts of Interest

None declared.

References

1. Bastow S, Dunleavy P, Tinkler J. The Impact of the Social Sciences: How Academics and Their Research Make a Difference. London: Sage Publications Ltd; 2014.
2. Schnitzler K, Davies N, Ross F, Harris R. Using Twitter™ to drive research impact: a discussion of strategies, opportunities and challenges. Int J Nurs Stud 2016 Jul;59:15-26. [doi: 10.1016/j.ijnurstu.2016.02.004] [Medline: 27222446]
3. Sinnenberg L, Buttenheim AM, Padrez K, Mancheno C, Ungar L, Merchant RM. Twitter as a tool for health research: a systematic review. Am J Public Health 2017 Jan;107(1):e1-e8. [doi: 10.2105/AJPH.2016.303512] [Medline: 27854532]
4. Harris JK, Mansour R, Choucair B, Olson J, Nissen C, Bhatt J, Centers for Disease Control and Prevention. Health department use of social media to identify foodborne illness - Chicago, Illinois, 2013-2014. MMWR Morb Mortal Wkly Rep 2014 Aug 15;63(32):681-685 [FREE Full text] [Medline: 25121710]
5. Yom-Tov E, Borsa D, Cox IJ, McKendry RA. Detecting disease outbreaks in mass gatherings using internet data. J Med Internet Res 2014 Jun 18;16(6):e154 [FREE Full text] [doi: 10.2196/jmir.3156] [Medline: 24943128]
6. McNeill A, Harris PR, Briggs P. Twitter influence on UK vaccination and antiviral uptake during the 2009 H1N1 pandemic. Front Public Health 2016 Feb;4:26 [FREE Full text] [doi: 10.3389/fpubh.2016.00026] [Medline: 26942174]
7. Lee JL, DeCamp M, Dredze M, Chisolm MS, Berger ZD. What are health-related users tweeting? A qualitative content analysis of health-related users and their messages on twitter. J Med Internet Res 2014 Oct 15;16(10):e237 [FREE Full text] [doi: 10.2196/jmir.3765] [Medline: 25591063]
8. Yuan P, Bare MG, Johnson MO, Saberi P. Using online social media for recruitment of human immunodeficiency virus-positive participants: a cross-sectional survey. J Med Internet Res 2014 May 01;16(5):e117 [FREE Full text] [doi: 10.2196/jmir.3229] [Medline: 24784982]
9. Fenner Y, Garland S, Moore EE, Jayasinghe Y, Fletcher A, Tabrizi SN. Web-based recruiting for health research using a social networking site: an exploratory study. J Med Internet Res 2012;14(1):e20. [Medline: 22297093]
10. O'Connor A, Jackson L, Goldsmith L, Skirton H. Can I get a retweet please? Health research recruitment and the Twittersphere. J Adv Nurs 2014 Mar;70(3):599-609. [doi: 10.1111/jan.12222] [Medline: 23909740]
11. Kneafsey R, Brown S, Sein K, Chamley C, Parsons J. A qualitative study of key stakeholders' perspectives on compassion in healthcare and the development of a framework for compassionate interpersonal relations. J Clin Nurs 2016 Jan;25(1-2):70-79. [doi: 10.1111/jocn.12964] [Medline: 26403949]
12. Larson EB, Yao X. Clinical empathy as emotional labor in the patient-physician relationship. J Am Med Assoc 2005 Mar 02;293(9):1100-1106. [doi: 10.1001/jama.293.9.1100] [Medline: 15741532]
13. de Zulueta PC. Developing compassionate leadership in health care: an integrative review. J Healthc Leadersh 2016;8:1-10 [FREE Full text] [doi: 10.2147/JHL.S93724] [Medline: 29355200]

14. Maben J, Peccei R, Adams M, Robert G, Richardson A, Murrels T, et al. NIHR Evaluation, Trials and Studies Coordinating Centre. 2012. Exploring the relationship between patients' experience of care and the influence of staff motivation, affect and wellbeing Final Report URL: http://www.netscc.ac.uk/hsdr/files/project/SDO_A1_08-1819-213_V01.pdf [accessed 2018-03-15] [WebCite Cache ID 6xwKlDpqB]
15. King N. Template analysis. In: Qualitative Methods and Analysis in Organizational Research: A Practical Guide. London: Sage Publications; 1998:118-134.
16. Raleigh VS, Hussey D, Seccombe I, Qi R. Do associations between staff and inpatient feedback have the potential for improving patient experience? An analysis of surveys in NHS acute trusts in England. Qual Saf Health Care 2009 Oct;18(5):347-354. [doi: 10.1136/qshc.2008.028910] [Medline: 19812096]
17. Maben J, Adams M, Peccei R, Murrells T, Robert G. 'Poppets and parcels': the links between staff experience of work and acutely ill older peoples' experience of hospital care. Int J Older People Nurs 2012 Jun;7(2):83-94 [FREE Full text] [doi: 10.1111/j.1748-3743.2012.00326.x] [Medline: 22531048]
18. Burnell L, Agan DL. Compassionate care: can it be defined and measured? The development of the Compassionate Care Assessment Tool (CCAT). IJCS 2013;6(2):180-187 [FREE Full text]
19. National Health Service. The National Archives. 2009. NHS Health and wellbeing: Final Report URL: http://webarchive.nationalarchives.gov.uk/20130124052412/http://www.dh.gov.uk/prod_consum_dh/groups/dh_digitalassets/documents/digitalasset/dh_108907.pdf [WebCite Cache ID 6xwLehnv3]
20. Maben J, Shuldham C, Taylor C, Foot C, Dawson J. National Institute for Health Research. 2017. A longitudinal national evaluation of Schwartz Centre Rounds®: an intervention to enhance compassion in relationships between staff and patients through providing support for staff and promoting their wellbeing URL: https://www.journalslibrary.nihr.ac.uk/programmes/hsdr/130749/%23/ [WebCite Cache ID 6xwLmZSNy]
21. Ebert JF, Huibers L, Christensen B, Christensen MB. Paper- or web-based questionnaire invitations as a method for data collection: cross-sectional comparative study of differences in response rate, completeness of data, and financial cost. J Med Internet Res 2018;20(1):e24 [FREE Full text] [doi: 10.2196/jmir.8353]
22. Scott A, Jeon SH, Joyce CM, Humphreys JS, Kalb G, Witt J, et al. A randomised trial and economic evaluation of the effect of response mode on response rate, response bias, and item non-response in a survey of doctors. BMC Med Res Methodol 2011 Sep 05;11:126 [FREE Full text] [doi: 10.1186/1471-2288-11-126] [Medline: 21888678]
23. Chai PR, Rosen RK, Lewis DM, Ranney ML, Boyer EW. Crowd-sourced focus groups on twitter: 140 characters of research insight. 2016 Presented at: Proceedings of the 50th Hawaii International Conference on System Sciences; January 2017; Hawaii p. 3746-3753 URL: https://www.researchgate.net/publication/317331302_Crowd-Sourced_Focus_Groups_on_Twitter_140_Characters_of_Research_Insight [doi: 10.24251/HICSS.2017.453]
24. British Psychological Society. 2013. Ethics guidelines for internet-mediated research URL: https://www1.bps.org.uk/system/files/Public%20files/Ethics%20Guidelines%20for%20Internet-mediated%20Research%202017%20Revision%20WEB.pdf [accessed 2018-03-15] [WebCite Cache ID 6xwMWMkeR]

Abbreviations

NHS: National Health Service

Edited by G Eysenbach; submitted 13.03.17; peer-reviewed by Y Sawbridge, KH Chu; comments to author 10.08.17; revised version received 24.11.17; accepted 28.02.18; published 23.04.18.

Please cite as:
Clyne W, Pezaro S, Deeny K, Kneafsey R
Using Social Media to Generate and Collect Primary Data: The #ShowsWorkplaceCompassion Twitter Research Campaign
JMIR Public Health Surveill 2018;4(2):e41
URL: http://publichealth.jmir.org/2018/2/e41/
doi:10.2196/publichealth.7686
PMID:29685866

Original Paper

The Development of a Social Networking–Based Relatedness Intervention Among Young, First-Time Blood Donors: Pilot Study

Victoria Frye[1,2], MPH, DrPH; Louisa Duffy[2], BA; Janis L France[3], PhD; Debra A Kessler[2], RN, MSc; Mark Rebosa[2], MBA; Beth H Shaz[2], MD; Bruce W Carlson[3], PhD; Christopher R France[3], PhD

[1]City University of New York School of Medicine, Community Health and Social Medicine, City University of New York, New York, NY, United States

[2]New York Blood Center, New York, NY, United States

[3]Ohio University, Department of Psychology, Athens, OH, United States

Corresponding Author:
Christopher R France, PhD
Ohio University
Department of Psychology
251 Porter Hall
Ohio University
Athens, OH, 45701
United States
Phone: 1 740 593 1079
Fax: 1 740 593 0579
Email: France@ohio.edu

Abstract

Background: Increasing repeat blood donation behavior is a critical public health goal. According to self-determination theory, the process of developing internal motivation to give blood and an associated self-identity as a blood donor may be promoted by feelings of “relatedness” or a connection to other donors, which may be enhanced through social relations and interactions.

Objective: The purpose of this report it to describe the development and pilot testing of a social networking-based (Facebook) intervention condition designed to increase feelings of relatedness via virtual social interaction and support.

Methods: To develop the intervention condition content, images, text, polls, and video content were assembled. Ohio University college students (N=127) rated the content (82 images/text) presented by computer in random order using a scale of one to five on various dimensions of relatedness. Mean ratings were calculated and analyses of variance were conducted to assess associations among the dimensions. Based on these results, the relatedness intervention was adapted and evaluated for feasibility, acceptability, and preliminary efficacy among 24 first-time donors, aged 18 to 24 years, in a 30-day pilot trial. Paired t-tests were conducted to examine change over time in relatedness and connectedness.

Results: The intervention condition that was developed was acceptable and feasible. Results of the uncontrolled, preintervention, and postintervention evaluation revealed that feelings of individual-level relatedness increased significantly after the intervention.

Conclusions: By promoting first-time blood donor relatedness, our goal is to enhance internal motivation for donating and the integration of the blood donor identity, thus increasing the likelihood of future repeat donation.

Trial Registration: ClinicalTrials.gov NCT02717338; https://clinicaltrials.gov/ct2/show/NCT02717338 (Archived by WebCite at http://www.webcitation.org/6ymHRBCwu)

(JMIR Public Health Surveill 2018;4(2):e44) doi:10.2196/publichealth.8972

KEYWORDS

blood donation; social media intervention; donor identity; internal motivation

Introduction

For decades, research has been conducted to understand repeat blood donation, with the goal of increasing donation efficiency, retention, and diversity of the donor pool. With less than 5% of the US population donating the entire national blood supply, increasing both numbers of blood donors and consistency of donation behavior is critical [1]. Novel approaches and methods are needed to cultivate and grow the core of the blood donor pool: the repeat donor [2]. Repeat donors continue to represent a small group of individuals responsible for donating a significant proportion of the blood in the United States [3]. Repeat donors are particularly critical to the blood supply because they are consistent, committed, and can be counted on to donate in emergencies and under suboptimal conditions. These individuals also tend to be a safer source of blood, for a variety of reasons, and contribute to the overall safety of the blood supply. Finally, in addition to transfusion services needing an appropriate amount of blood, the blood received needs to match the patients' needs, particularly for sickle cell disease patients. Repeat donors of select blood types and characteristics help transfusion services better meet the clinical needs of the institution's population, resulting in improved clinical outcomes.

The multi-site, longitudinal National Heart, Lung, and Blood Institute Retrovirus Epidemiology Donor Study (REDS) has followed donors since 1989 and has found four general groups of blood donors: established or multi-year repeat donors; single-year repeat donors; first-time donors; and donors who donated once but never again [4]. Studies of repeat donors have found that they are more often male, white, and with higher education levels [5]; a more recent analysis of donors in a large metropolitan area also found that repeat donors were more likely to be male, white, and older [6]. Although analyses of sociodemographic characteristics are of value, identifying malleable characteristics that act to increase repeat donation behavior is needed to craft effective interventions. Numerous studies have found attitudes, altruism, feelings of social responsibility, self-esteem, self-efficacy, and social pressure to be related to blood donation intention and behavior [7-14]. Donors who give blood more than once within a year of their first donation are associated with long-term donation [15]. Repeat donation varies among different groups of donors; for new donors the intention to donate and age may predict repeat donation behavior, whereas for repeat donors, anticipated regret and a "moral norm" (or feeling of responsibility or duty) may be stronger predictors of repeat donation behavior [16]. Finally, motivation to donate has been found to shift over time for approximately one third of repeat donors, with feelings of solidarity and duty increasing in importance [17]; this suggests the potential that group-based norms and connectedness may influence repeat donor behavior.

Charng and colleagues [18], in their germinal study of donation intentions and behavior, focused attention on social relations and identity as predictors of repeat donation behavior. The authors framed identity as a sense of self that includes various "role identities" which may be established by behavior and reinforced through repetitive or habitual behavior [18]. The importance of the role identity to the sense of self is termed "salience" [19,20]; furthermore, role identity is that part of an individual's self-concept that derives from membership in a social role or group, along with the attitudes, values, and emotional significance attached to group affiliation [21]. Integrating a role identity as a blood donor into a sense of self may be triggered by an initial positive donation experience and reinforced through repeat donation, which may eventually become habitual. Categorizing oneself as a member of a social group or role may precede the behavior and subsequent repeated behavior [22]. The influence of the social group with which the donor affiliates, and where group members share a group or role identity and engage in social interaction, impacts repeated behavior. Group membership may encourage and reinforce specific behaviors through social interaction, support, connectedness or relatedness, and solidarity [19-21].

A social relations/role identity approach to integrating an identity as a blood donor by increasing feelings of relatedness or connectedness to other donors is also consistent with self-determination theory. Integrating the "blood donor" role into an individual's identity enhances intrinsic motivation, where autonomous, intrinsically motivated behavior springs from inherent satisfaction/commitment and/or interest [23]. Few interventions have attempted to enhance self-concept or identity as a blood donor by increasing the feeling of connectedness or relatedness to fellow donors through social interaction. Social networking platforms may be ideal vehicles to help develop such interactions and to foster the feelings of relatedness that can encourage donor identity formation. Here we describe the development and assessment of the feasibility, acceptability, and preliminary efficacy of a social network-based (Facebook) intervention condition designed to increase feelings of relatedness among young, first-time blood donors in New York City. This Facebook-based intervention condition was developed and tested in preparation for inclusion as one of four conditions in an ongoing, eight-group, full factorial design, randomized controlled trial of a multi-component intervention to increase blood donation among first-time donors [24].

Methods

The relatedness intervention condition was developed as a "secret" Facebook group for first-time blood donors. First, we describe the methods used to develop and evaluate the content for the Facebook group; next, we describe methods used to evaluate the preliminary efficacy of the intervention condition on measures of relatedness and social network connectedness. This research was approved by the Institutional Review Boards of New York Blood Center (NYBC) and Ohio University, and registered at ClinicalTrials.gov (NCT02717338).

Facebook Relatedness Intervention Content Development

To develop the relatedness intervention condition content, we identified images, text, polls, and videos used in previous NYBC social media efforts; novel content, in the form of images and text, was also developed. Each image or text was selected and categorized by the first and second author according to whether they were perceived to increase donor feelings of being supported, respected, valued, connected/belonging, inspired to donate again, or having shared values with other donors. In order to develop the final set of images and text for the intervention, we conducted an evaluation among the focal population of whether the content inspired feelings of relatedness, specifically: belonging to a group, connectedness to other donors, shared values with other donors, and feeling respected by others. Ohio University college students (N=127) rated the content (82 images/text) using a scale of one to five

(1 = least positive; 5 = most positive) on the dimensions described above, as well as how much the image was "liked" and was "cool." Images/text were presented to students via computer and in random order to control for order effects. Mean ratings for each image/text were calculated; analysis of variance (ANOVA) was conducted to assess associations among the rating dimensions. Based on these results, not all of the original 82 images were included in the pilot and some were adapted for use in the pilot study. Highly rated images were included and novel content, designed to encourage social interaction around donation experiences (eg, polls about specific aspects of the most recent donation experience or friends and family who donate), positive identity formation (posts with text/images related to "being a hero" or "saving a life" through donation), and group affiliation and connectedness (logos and visuals specific to the secret Facebook study group that could also be posted to participants' pages) were developed and included in the pilot trial of the Facebook intervention condition.

Facebook Relatedness Intervention Condition Pilot

To evaluate the condition's feasibility, acceptability, and preliminary efficacy, we recruited recent first-time NYBC donors, aged 18 to 24, who agreed to participate in a 30-day pilot study. Invitations to participate were emailed to 95 donors and resulted in an initial recruitment of 24 donors (ie, a response rate of 25%). Donors who agreed to participate completed an online consent form and were invited to join the group for 30 days; they were informed that they would be required to join a "new donor" secret Facebook group. All participants were existing Facebook users. This group was used to expose participants to blood donation-related content and to provide a forum for them to discuss their experiences with blood donation, post images of their donations, and/or describe reasons for donating. Prior to joining the group, participants completed a set of baseline (preintervention) surveys which included basic sociodemographic information and measures of social media connectedness and donor relatedness (described in detail below). To join the secret group, participants were instructed to "friend" the group administrator (the second author), who then sent them an invitation to join the group. Although only current secret group members were able to see posts from the group [25], as an extra safeguard of confidentiality group participants were provided information on how to set their privacy settings prior to joining the group. Participants received a US $30 check if they provided informed consent to participate (n=24), completed the predonation survey (n=24), and joined the Facebook group and completed the post-intervention survey (n=18). The secret group ran for 30 days (SD 7 days) for each participant and was moderated by the second author, who posted content daily and ensured that participant postings conformed to the NYBC policy on acceptable use of social media. Two posts were added each day, with each post typically including an image and text that acted as an invitation for group members to answer a question or poll in an effort to encourage interaction. Once the participant spent 30 days in the group, they were removed from the group and a link to an online follow-up (postintervention) survey was sent to them via email. Paired t-tests were conducted to examine change over time in relatedness and connectedness (see below for a description of the development of the outcome measures). Finally, we assessed acceptability of the Facebook group with three closed-ended items (using true/false response options) that assessed whether participants thought the Facebook group was a good way to connect to other donors or made them want to donate again. We also asked if the participants wished they could have stayed a part of the Facebook group and whether they created a new Facebook account to participate. Finally, one open-ended item was included to generate feedback on how to improve the Facebook group. Responses to this open-ended item were brief and direct, allowing us to summarize the content across all 18 participants.

Measures

No validated measures existed to assess relatedness or social network connectedness among young blood donors, so we adapted existing measures for this pilot. The donor relatedness measure (Multimedia Appendix 1) was adapted from several sources [26-28]. From the Vlachopoulos and Michailidou scale [28], which was designed to assess relatedness as a subscale of a larger scale to measure autonomy, competence, and relatedness in exercise, we adapted two items reflecting feelings of comfort and friendliness with other blood donors. From the Furrer and Skinner scale [26], originally designed to measure relatedness in a school setting, we adapted one item related to feeling accepted. From Sheldon and Hilpert's scale [27], originally designed to assess relatedness among Facebook users, we adapted six items reflecting feelings of intimacy and appreciation, as well as aloneness. These nine adapted items were originally included based on their centrality to the construct. We included a single team-developed item, "I feel like a part of a blood donor community." These 10 items were administered to 923 Ohio University college students with a history of blood donation, using a 7-point response scale (1=not at all; 7=extremely). A principal component factor analysis with varimax rotation was run on the original 10-item scale. Results of this analysis supported a 9-item scale, accounting for 81% of the variance, with items loading on three factors: group-level relatedness (Cronbach α=0.98), individual-level relatedness (Cronbach α=0.86), and nonrelatedness (Cronbach α=0.65). Group-level and individual-level relatedness were positively correlated (r=0.38). Sample items included, "I feel like part of the blood donor community" (group-level relatedness), "I feel a strong sense of intimacy with other blood donors" (individual-level relatedness), and, "I feel alone as a blood donor" (non-relatedness). The social media connectedness scale (Multimedia Appendix 2) was adapted from Grieve et al [29] and tested in the same college student sample. Respondents again rated each item using a 7-point response scale (1=not at all; 7=extremely). Results of a principal components factor analysis varimax rotation supported a final 12-item scale, accounting for 60% of the variance, with items loading on two factors: social media connectedness (Cronbach α=0.87) and social media disconnectedness (Cronbach α=0.85). These scales were negatively correlated (r=-0.42). Sample items included, "I feel close to people on social media" and, "I don't feel related to most people on social media."

Results

Evaluation of the Relatedness Intervention Content

The student sample (N=127) self-identified as 55% (70/124) women and 45% men (57/127); 42% donors (53/127) and 58% nondonors (74/124). For 80 of the images (80/82, 98%) the “like” ratings did not differ significantly as a function of donor status. As shown in Figure 1, when the images were divided into four groups based on their “like” ratings, ANOVAs revealed that the most liked images were also associated with the most positive responses across each of the other seven domains.

Examination of the content of the most liked versus least liked images revealed a distinct preference for pictures that conveyed a sense of connection with other people or a direct contribution to the welfare of others. Figure 2 and Figure 3 depict the top ten “most liked” images, the theoretical targets on the various images, as well as the average student ratings by “likes,” “inspired to donate,” and “connectedness.” Student ratings of the same image varied by theoretical target (eg, inspired to donate vs connectedness) although ratings did not vary by donor versus nondonor status. This information informed the selection of images for use in the pilot trial of the relatedness intervention condition.

Relatedness Condition Intervention Pilot

Of the 24 participants who consented to the study, 75% (18/24) agreed to participate in the relatedness intervention condition pilot trial and completed both the preintervention and postintervention assessments. Interaction with the content was limited, with most posts “liked” by just a few pilot participants. Nonetheless, our analysis revealed a significant increase in feelings of individual-level relatedness from preintervention (mean=3.1, SD=1.2) to postintervention (mean=4.8, SD=1.3; t(17)=5.32, *P*<.001). No significant change was observed for group-level relatedness (*P*=.45) or nonrelatedness (*P*=.64) subscales. There was a marginal increase in social media connectedness from preintervention (mean=3.9, SD=1.3) to postintervention (mean=4.2, SD=1.5; t(17)=1.82, *P*=.09), but no significant change in social media disconnectedness (*P*=.71). We evaluated acceptability using a postintervention assessment. Over three-quarters of participants (14/18, 78%) thought the Facebook group was a good way to connect to other donors and the majority (16/18, 89%) reported that it made them want to donate again. Over half of the respondents (10/18, 56%) reported that they wished they could have stayed a part of the Facebook group. Suggestions to improve the Facebook group included: providing more concrete information on how to donate, actively encouraging group members to share and post, personal messaging to participants, allowing group chats, adding more interactive polls and questions, and adding more participants to the group.

Figure 1. Ratings of images by “like” quartile (N=127).

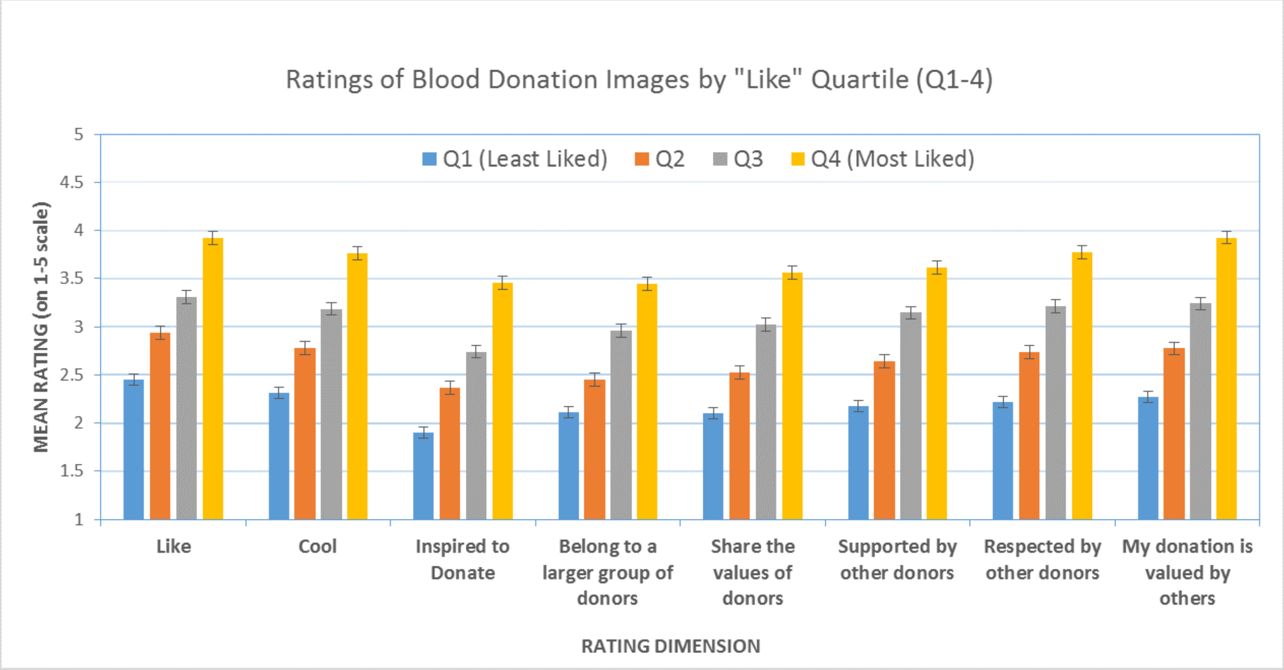


Figure 2. Top five most "liked" images, Facebook relatedness content evauation (N=127).

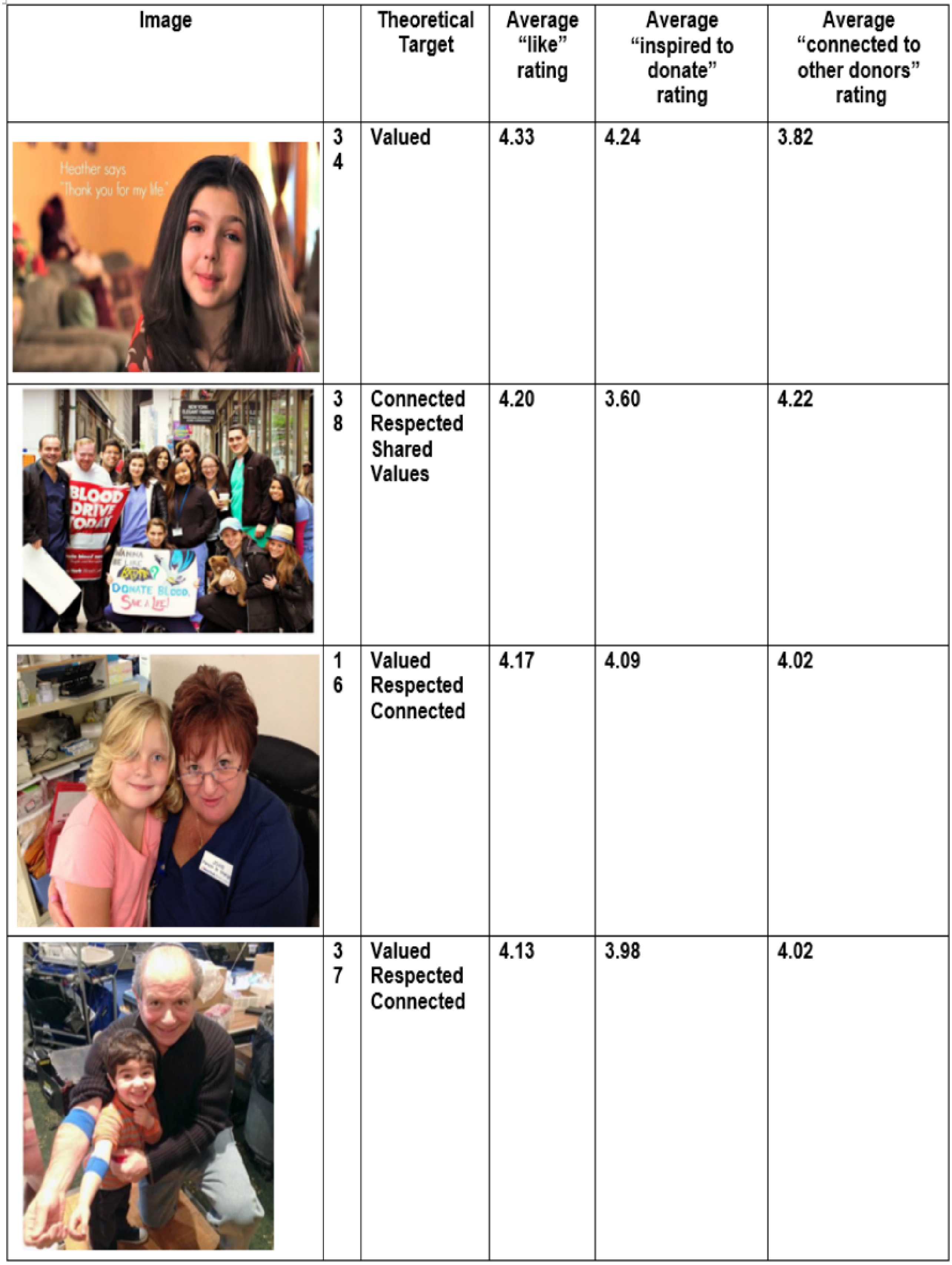
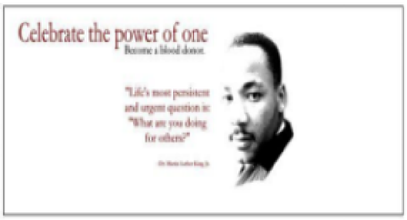
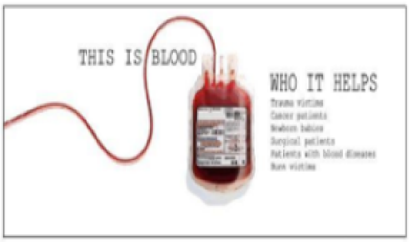


Image		Theoretical Target	Average "like" rating	Average "inspired to donate" rating	Average "connected to other donors" rating
Heather says "Thank you for my life"	34	Valued	4.33	4.24	3.82
	38	Connected Respected Shared Values	4.20	3.60	4.22
	16	Valued Respected Connected	4.17	4.09	4.02
	37	Valued Respected Connected	4.13	3.98	4.02

Figure 3. Next five most "liked" images, Facebook relatedness content evauation (N=127).

	21	Inspired to donate Shared Values	4.02	3.93	3.69
	45	Inspired to donate Valued Connected	4.00	3.96	3.64
	31	Inspired to donate Valued	3.98	3.78	3.60
	22	Inspired to donate Valued	3.96	3.74	3.65
	41	Valued Connected	3.93	3.67	3.57

Discussion

Principal Findings

In our development and pilot testing of a social network-based relatedness intervention condition for young, first-time blood donors we found that the 30-day membership in a secret and monitored Facebook group was acceptable and feasible. No donors had to create a new account to participate and most continued participation through the end of the 30-day pilot period. In addition, although we did not detect a high volume of interaction among the participants or with the content posted, our pre-post analyses revealed a significant increase in self-reported feelings of blood donor relatedness and a marginally significant increase in social media connectedness. As posited by self-determination theory, an intervention that promotes a sense of belonging or connectedness may aid in satisfying the fundamental human need for relatedness and contribute to a more internalized and self-determined motivation regarding the behavior in question [30]. Thus, we anticipate that this intervention may foster a greater intrinsic motivation regarding blood donation and promote the likelihood of repeat donation behavior.

It is important to note that although we anticipated a postintervention increase in both group- and individual-level relatedness, a significant change was seen only for the individual-level relatedness subscale. Group-level relatedness items tapped feelings of acceptance, comfort, and friendliness with other donors as well as being a part of the blood donor community. Individual-level items reflected feelings of connectedness, intimacy, and contact. Post hoc analyses revealed that a significant change was seen for each item from the individual-level relatedness scale, while no items on the group-level relatedness scale significantly differed from preintervention to postintervention. We may have seen this difference because the social network platform-based intervention acted specifically to increase contact among members thus increasing “individual-level” feelings of relatedness. It is possible that the feelings tapped in the group-level relatedness subscale may change more gradually in response to the intervention than those measured on the individual-level subscale. For instance, if a certain level of individual relatedness is necessary for group-level relatedness to emerge or increase, then we would expect to see a time lag between increases in individual relatedness and group relatedness. If so, the postintervention assessment may have occurred too soon for any lagged group-level relatedness changes to be observed. Further investigation into the observed discrepancy between the intervention effects on group- versus individual-level relatedness is warranted. However, it must also be noted that this is an uncontrolled pilot of a condition and thus caution is urged in interpreting these results.

Another question we will consider in future research using this intervention is whether the level of social interaction in the Facebook group influences the magnitude of change in both individual- and group-level relatedness between participants. As noted, there was limited interaction with the media posted to the secret group page, and yet a significant effect was still seen for individual-level relatedness, along with a marginally significant increase in social media connectedness. A larger sample size, along with the collection of Facebook metrics that were not available in this pilot study, will allow us to evaluate whether a user’s level of engagement with the social network intervention exhibits a dose-response effect on donor relatedness and connectedness.

Limitations

Although this pilot study provides initial support for the feasibility and acceptability of using a social network platform to increase relatedness among new blood donors, as with any new intervention, this preliminary research has limitations. A potential limitation is that the posts (images and text) were tested among a sample of students in Ohio, whereas the pilot was conducted with donors living in the New York City area. Another potential limitation relates to the lack of a control condition in this pilot trial. In the absence of a suitable control group, it is possible that the positive changes in relatedness were due to demand characteristics and socially desirable responses from participants. Interestingly, the fact that we did not observe the anticipated positive changes in group-level relatedness may be seen as support for the notion that the observed significant effects in individual-level relatedness are not merely an indication of socially desirable responding, as we would expect such an effect across all measures.

An additional limitation regarding the potential efficacy of this intervention relates to whether we can assume that the findings based on a group of donors who volunteer to participate in a Facebook-based study are generalizable to the larger population of donors. Whereas not all donors may be interested or willing to participate in a social network-based intervention, it is reasonable to assume that those who do so may also be the most likely to engage with and benefit from such an approach. Finally, some limitations exist regarding the measures used in this pilot study. While preliminary psychometric testing was conducted in designing our measures of donor relatedness and social connectedness, these scales were developed using a sample of Midwestern college students. Further validation using groups of individuals with greater diversity in age, donation experience, ethnicity, and racial backgrounds is warranted. In addition, as previously mentioned, Facebook metrics that may allow us to gather more data about how participants are interacting with the secret group, and how this may impact relatedness outcomes, were not available in this pilot study. However, this information will be examined as part of an ongoing randomized controlled trial that applies Social Determination Theory to the blood donation context in an effort to enhance retention of first-time blood donors using a combination of donor competence, autonomy, and relatedness interventions [24].

Conclusions

This pilot study developed and tested the feasibility and preliminary efficacy of a social network platform-based intervention designed to increase a sense of blood donor relatedness among group members. Results of this small, uncontrolled trial suggest that a monitored Facebook group devoted to promoting a blood donor identity may be an effective tool for increasing a feeling of donor relatedness and connectedness among young donors.

Acknowledgments

All authors contributed to the conceptualization, writing, and editing of this report. Research reported in this publication was supported by the National Heart, Lung, and Blood Institute of the National Institutes of Health under Award Number R01HL127766. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Blood Donor Relatedness Scale.

[PDF File (Adobe PDF File), 32KB - publichealth_v4i2e44_app1.pdf]

Multimedia Appendix 2

Social Media Connectedness Scale.

[PDF File (Adobe PDF File), 37KB - publichealth_v4i2e44_app2.pdf]

References

1. Whitaker B, Henry R. Department of Health and Human Services. Washington, DC; 2013. The 2011 national blood collection and utilization survey report URL: https://www.aabb.org/research/hemovigilance/bloodsurvey/Documents/11-nbcus-report.pdf [accessed 2018-04-11] [WebCite Cache ID 6ybPuUthv]
2. Gillespie TW, Hillyer CD. Blood donors and factors impacting the blood donation decision. Transfus Med Rev 2002 Apr;16(2):115-130. [doi: 10.1053/tmrv.2002.31461] [Medline: 11941574]
3. Schreiber GB, Sanchez AM, Glynn SA, Wright DJ, Retrovirus Epidemiology Donor Study. Increasing blood availability by changing donation patterns. Transfusion 2003 May;43(5):591-597. [Medline: 12702179]
4. Zuck TF, Thomson RA, Schreiber GB, Gilcher RO, Kleinman SH, Murphy EL, et al. The Retrovirus Epidemiology Donor Study (REDS): rationale and methods. Transfusion 1995;35(11):944-951. [Medline: 8604493]
5. Murphy EL, Shaz B, Hillyer CD, Carey P, Custer BS, Hirschler N, NHLBI Retrovirus Epidemiology in Blood Donors Study-II (REDS-II). Minority and foreign-born representation among US blood donors: demographics and donation frequency for 2006. Transfusion 2009 Oct;49(10):2221-2228 [FREE Full text] [doi: 10.1111/j.1537-2995.2009.02271.x] [Medline: 19555415]
6. Shaz BH, James AB, Hillyer KL, Schreiber GB, Hillyer CD. Demographic patterns of blood donors and donations in a large metropolitan area. J Natl Med Assoc 2011 Apr;103(4):351-357. [Medline: 21805814]
7. Sanchez AM, Ameti DI, Schreiber GB, Thomson RA, Lo A, Bethel J, Retrovirus Epidemiology Donor Study. The potential impact of incentives on future blood donation behavior. Transfusion 2001 Feb;41(2):172-178. [Medline: 11239218]
8. France J, France C, Himawan L. A path analysis of intention to redonate among experienced blood donors: an extension of the theory of planned behavior. Transfusion 2007 Jun;47(6):1006-1013. [doi: 10.1111/j.1537-2995.2007.01236.x] [Medline: 17524090]
9. France JL, France CR, Himawan LK. Re-donation intentions among experienced blood donors: does gender make a difference? Transfus Apher Sci 2008 Apr;38(2):159-166. [doi: 10.1016/j.transci.2008.01.001] [Medline: 18331813]
10. Fernández Montoya A, de Dios Luna del Castillo J, López Berrio A, Rodríguez Fernández A. [Attitudes, beliefs, and motivations in blood donors and non-donors]. Sangre (Barc) 1996 Dec;41(6):427-440. [Medline: 9148420]
11. Moore R. Promoting blood donation: a study of the social profile, attitudes, motivation and experience of donors. Transfus Med 1991 Dec;1(4):201-207. [Medline: 9259850]
12. Steele WR, Schreiber GB, Guiltinan A, Nass C, Glynn SA, Wright DJ, Retrovirus Epidemiology Donor Study. The role of altruistic behavior, empathetic concern, and social responsibility motivation in blood donation behavior. Transfusion 2008 Jan;48(1):43-54. [doi: 10.1111/j.1537-2995.2007.01481.x] [Medline: 17894795]
13. Sojka BN, Sojka P. The blood donation experience: self-reported motives and obstacles for donating blood. Vox Sang 2008 Jan;94(1):56-63. [doi: 10.1111/j.1423-0410.2007.00990.x] [Medline: 18171329]
14. Giles M, McClenahan C, Cairns E, Mallet J. An application of the theory of planned behaviour to blood donation: the importance of self-efficacy. Health Educ Res 2004 Aug;19(4):380-391. [doi: 10.1093/her/cyg063] [Medline: 15155590]

15. Schreiber GB, Sharma UK, Wright DJ, Glynn SA, Ownby HE, Tu Y, Retrovirus Epidemiology Donor Study. First year donation patterns predict long-term commitment for first-time donors. Vox Sang 2005 Feb;88(2):114-121. [doi: 10.1111/j.1423-0410.2005.00593.x] [Medline: 15720609]
16. Godin G, Conner M, Sheeran P, Bélanger-Gravel A, Germain M. Determinants of repeated blood donation among new and experienced blood donors. Transfusion 2007 Sep;47(9):1607-1615. [doi: 10.1111/j.1537-2995.2007.01331.x] [Medline: 17725724]
17. Fernández-Montoya A, López-Berrio A, Luna del Castillo JD. How some attitudes, beliefs and motivations of Spanish blood donors evolve over time. Vox Sang 1998;74(3):140-147. [Medline: 9595640]
18. Charng H, Piliavin J, Callero P. Role identity and reasoned action in the prediction of repeated behavior. Soc Psychol Q 1988 Dec;51(4):303-317. [doi: 10.2307/2786758]
19. Stryker S. Identity salience and role performance: the relevance of symbolic interaction theory for family research. J Marriage Fam 1968 Nov;30(4):558-564. [doi: 10.2307/349494]
20. Turner RH. The role and the person. Am J Sociol 1978 Jul;84(1):1-23. [doi: 10.1086/226738]
21. Tajfel H. Differentiation Between Social Groups: Studies in the Social Psychology of Intergroup Relations. London: Academic Press; 1978:A.
22. Stets J, Burke P. Identity theory and social identity theory. Soc Psychol Q 2000 Sep;63(3):224-237. [doi: 10.2307/2695870]
23. Deci E, Ryan R. The “what” and “why” of goal pursuits: human needs and the self-determination of behavior. Psychol Inq 2000 Oct;11(4):227-268. [doi: 10.1207/S15327965PLI1104_01]
24. France CR, France JL, Carlson BW, Frye V, Duffy L, Kessler DA, et al. Applying self-determination theory to the blood donation context: the blood donor competence, autonomy, and relatedness enhancement (Blood Donor CARE) trial. Contemp Clin Trials 2017 Dec;53:44-51 [FREE Full text] [doi: 10.1016/j.cct.2016.12.010] [Medline: 27979752]
25. Facebook. 2017 Aug 17. What are the privacy settings for groups? URL: https://www.facebook.com/help/220336891328465 [accessed 2018-04-11] [WebCite Cache ID 6ybRrfJcd]
26. Furrer C, Skinner E. Sense of relatedness as a factor in children's academic engagement and performance. J Educ Psychol 2003;95(1):148-162. [doi: 10.1037/0022-0663.95.1.148]
27. Sheldon K, Hilpert J. The balanced measure of psychological needs (BMPN) scale: an alternative domain general measure of need satisfaction. Motiv Emot 2012 Mar 3;36(4):439-451. [doi: 10.1007/s11031-012-9279-4]
28. Vlachopoulos S, Michailidou S. Development and initial validation of a measure of autonomy, competence, and relatedness in exercise: the Basic Psychological Needs in Exercise Scale. Meas Phys Educ Exerc Sci 2006 Sep;10(3):179-201. [doi: 10.1207/s15327841mpee1003_4]
29. Grieve R, Indian M, Witteveen K, Anne Tolan G, Marrington J. Face-to-face or Facebook: can social connectedness be derived online? Comput Human Behav 2013 May;29(3):604-609. [doi: 10.1016/j.chb.2012.11.017]
30. Ryan RM, Deci EL. Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. American Psychologist 2000 Jan;55(1):68-78. [doi: 10.1037/0003-066X.55.1.68] [Medline: 11392867]

Abbreviations

ANOVA: analysis of variance
NYBC: New York Blood Center

Edited by G Eysenbach; submitted 15.09.17; peer-reviewed by S Bhatlekar, H Degenholtz, M Puttaroo; comments to author 03.12.17; revised version received 28.01.18; accepted 15.02.18; published 26.04.18.

Please cite as:
Frye V, Duffy L, France JL, Kessler DA, Rebosa M, Shaz BH, Carlson BW, France CR
The Development of a Social Networking–Based Relatedness Intervention Among Young, First-Time Blood Donors: Pilot Study
JMIR Public Health Surveill 2018;4(2):e44
URL: http://publichealth.jmir.org/2018/2/e44/
doi:10.2196/publichealth.8972
PMID:29699961

Original Paper

Junk Food Marketing on Instagram: Content Analysis

Amy Jo Vassallo[1], MCHD; Bridget Kelly[2], PhD; Lelin Zhang[3], PhD; Zhiyong Wang[3], PhD; Sarah Young[1], MPH; Becky Freeman[1], PhD

[1]Prevention Research Collaboration, School of Public Health, Charles Perkins Centre, The University of Sydney, Sydney, NSW, Australia
[2]Early Start Research Institute, School of Health and Society, University of Wollongong, Sydney, NSW, Australia
[3]School of Information Technologies, The University of Sydney, Sydney, NSW, Australia

Corresponding Author:
Amy Jo Vassallo, MCHD
Prevention Research Collaboration, School of Public Health
Charles Perkins Centre
The University of Sydney
Level 6 (D17), Charles Perkins Centre
The University of Sydney
Sydney, NSW, 2006
Australia
Phone: 61 2 8627 4449
Email: amy.vassallo@sydney.edu.au

Abstract

Background: Omnipresent marketing of processed foods is a key driver of dietary choices and brand loyalty. Market data indicate a shift in food marketing expenditures to digital media, including social media. These platforms have greater potential to influence young people, given their unique peer-to-peer transmission and youths' susceptibility to social pressures.

Objective: The aim of this study was to investigate the frequency of images and videos posted by the most popular, energy-dense, nutrient-poor food and beverage brands on Instagram and the marketing strategies used in these images, including any healthy choice claims.

Methods: A content analysis of 15 accounts was conducted, using 12 months of Instagram posts from March 15, 2015, to March 15, 2016. A pre-established hierarchical coding guide was used to identify the primary marketing strategy of each post.

Results: Each brand used 6 to 11 different marketing strategies in their Instagram accounts; however, they often adhered to an overall theme such as athleticism or relatable consumers. There was a high level of branding, although not necessarily product information on all accounts, and there were very few health claims.

Conclusions: Brands are using social media platforms such as Instagram to market their products to a growing number of consumers, using a high frequency of targeted and curated posts that manipulate consumer emotions rather than present information about their products. Policy action is needed that better reflects the current media environment. Public health bodies also need to engage with emerging media platforms and develop compelling social counter-marketing campaigns.

(JMIR Public Health Surveill 2018;4(2):e54) doi:10.2196/publichealth.9594

KEYWORDS

food and beverage; advertisements; social media

Introduction

Background

One of the most significant and preventable causes of poor health and early death is the rapidly rising rates of overweight and obesity. Excess adiposity is a major risk factor for a number of chronic conditions, including cardiovascular disease, type 2 diabetes, osteoarthritis, and some cancers [1,2]. In 2010 alone, it was estimated that, globally, overweight and obesity caused 3.4 million deaths [3].

The alarming increase in obesity prevalence worldwide [3] is predominantly driven by dietary changes, with food now increasingly processed, affordable, and aggressively marketed [2]. The majority of food consumed in developed countries is manufactured and distributed by large for-profit companies [4]. A key driver of company food product sales is highly effective and omnipresent marketing. Current evidence demonstrates that acute exposure to food advertising influences dietary choices [4,5] and increases food intake in children [6]. In adolescents and adults, advertising builds substantial brand awareness, particularly for fast food [7].

Young people's critical assessment of advertising requires the ability to discriminate between commercial and noncommercial content, recognition of the persuasive intent of advertising, and the ability to apply reasonable skepticism to it [8]. There is evidence that children younger than 8 years are not able to recognize the persuasive intent of advertising, and although these skills develop with time, the age children start to recognize these messages is unknown [8]. The advent of social media, which promotes engagement and friendship between brands and consumers, may diminish the ability to discriminate between commercial and noncommercial content [9].

With the rising popularity of social media, brands are using this platform to reach a diverse range of audiences, using a variety of powerful marketing strategies. Importantly and uniquely, social media allows for personalization of advertisements, which increases consumer receptivity of advertising, as it is tailored to their specific needs and interests [10]. Social media users can also communicate with their friends and followers about a brand or product, for example, by tagging a friend, reposting competitions, or tagging brands in their own personal posts. This extends the reach of individual advertisements and creates the sentiment that certain brands and products are endorsed by peers [10] or high-profile social media users, sportspeople, and celebrities [11].

Advertising has a strong influence on food preferences, choices, and consumption, particularly in children [6,8,12]. Adolescents are considered the primary targets for digital marketing [10] because of their own disposable income, spending power, ease of technology use, and role in setting and following trends [13]. Advertisers and food companies often assert that the purpose of their advertising is to provide information about their brands and products, including their healthier options, to allow consumers to make an informed choice about purchase and consumption [14]. It is, therefore, vital to gain a better understanding of how energy-dense, nutrient-poor (EDNP) food and beverages are being advertised on popular social media platforms and if the claims about the supposed purpose of advertising are justified.

Instagram is a popular mobile photo-sharing and, more recently, video-sharing social media platform. It was launched in October 2010, purchased by Facebook in 2012 [15], and in April 2017, there were more than 700 million monthly active users worldwide [16]. In 2015, Instagram was the most used social media platform among teenagers in the United States [17]. In 2016, it was the second most popular social media platform (after Facebook) among those aged older than 18 years [18] and more popular than Facebook in certain ethnic groups [19]. Many businesses have established their own accounts, as a growing number of people are using social media to "follow or find out about particular brands or businesses" [20]. In March 2017, Instagram reported that 8 million companies used the social media site—a rapid increase from the 1.5 million companies active on the platform in September 2016 [21].

Previous research has investigated how "junk" food and nonalcoholic beverages are promoted on Facebook [22] or the effects of user-generated images [23], and highlighted that a larger study is required to better understand the promotion elements used by brands on Instagram [24]. The engagement rate between advertiser and user is higher for Instagram than Facebook [25], indicating that, although fewer people have an Instagram account, there is potentially more total activity between brands and consumers. As Instagram was designed as a mobile app, it is frequently and repeatedly accessed. For example, globally, at least 60% of users log in daily [26], and in countries such as Australia, users check their accounts multiple times per day and an average of 26 times per week [20], which again demonstrates the potential for high reach of branded messages and normalize behavior [27].

Objectives

Due to the great potential of social media to influence the food and beverage choice of young people, it is essential to understand how brands are advertising their products online. Therefore, the primary aim of this study was to investigate the frequency of images and videos posted by the most popular ENDP food and beverages brands on Instagram and the marketing strategies used in these images. Secondarily, the study also aimed to evaluate if posts contained any product information, health claims, or healthier choice claims.

Methods

Selection of Instagram Accounts

A total of 15 of the most popular ENDP food and beverages accounts on Instagram were included. These were Ben and Jerry's, Burger King, Coca-Cola, Dominos, Gatorade, KFC, McDonald's, Monster Energy, Nutella, Oreo, Pepsi, Red Bull, Starbucks, Subway, and Taco Bell. Popular ENDP brands (excluding grocery stores) were first identified based on 2014 global sales rankings and checked for the presence of an Instagram account. If the account had at least 100,000 followers (in November 2015), it was included in the study (16 brands in total). For brands that maintained both global and nation-specific Instagram accounts, the global account was included. For each included brand, 12 months of Instagram posts from March 15, 2015, to March 15, 2016, were extracted, including the image, caption, number of likes and number of comments for each image, and for videos, the number of views. One brand (Wendy's) was then excluded from the study, as its first Instagram post was on July 10, 2015, which was partway through the study period.

Establishing the Coding Guide

A coding guide was used to identify the primary marketing strategy of each post. The first iteration of the guide was based on the categories developed in previously published works, including branded characters, branding elements, celebrities, children's characters, competitions, corporate social responsibility or philanthropy, engagement, links, special price promotions, sponsorships or partnerships, sportspeople, videos, and vouchers [22,28]. This guide was then used to code a subsample of 10 images from each Instagram account included in this study. If an image could not be suitably coded into any of the existing categories, a new category was added in an iterative process. This was done to ensure the coding frame was suitable for Instagram content. Two new marketing categories emerged: product imagery and image with no clear marketing strategy (which included images unrelated to the brand or product, for example, the view from an employee's window).

The coding guide also contained instructions to code each post regarding whether it contained content that was informational (yes or no), original (yes or no), or portrayed the products as healthy choices or contained health claims (yes or no). Table 1 contains a summarized version of coding guide categories and definitions.

For each post, an indicator of marketing strategy was recorded. Where more than one marketing strategy featured in a single post, strategies were coded according to the following hierarchy: Corporate Social Responsibility or Philanthropy>Celebrities or Sportspeople or Children's characters or Branded characters>Special price promotions or vouchers or competitions>Engagement>Sponsorships or Partnerships>Videos>Links>Branding>Product imagery. For example, if a post contained a video of a celebrity holding a branded product, it was coded as Celebrity. The primary rationale in establishing this hierarchy was to ensure consistency between coders, rather than to imply one marketing strategy is more important or meaningful than another. However, the strategies that were anticipated to be the most common (eg, branding) were placed at the bottom of the hierarchy in an attempt to capture the breadth of marketing strategies used on Instagram. The informational content, original and health claim categories were not hierarchical, and each post was also assessed against these criteria.

Coding the Images

Two authors (AJV and SY) were responsible for coding each image. First, one complete account (Ben and Jerry's) was coded by both coders independently. All codes assigned for all images were checked for agreement, and any discrepancies were discussed and agreed upon, incorporating the opinion of a third coder (BF) if required. All other accounts were then coded by one coder each (AJV or SY). If coders were unsure how to categorize a post, it was discussed and coded together, again incorporating a third coder (BF) if required. Descriptive comments about key features of each account were also recorded narratively during the coding process.

Table 1. Definitions of marketing strategies and coding criteria.

Category	Definition
Assess each post for	
Informational content	Recognizable information about what the product looks/tastes like, ingredients, or place of purchase
Original (brand-generated) content	Image is generated by the brand, or third party sponsored by brand, as opposed to consumer
Health claims	Specific reference that the product shown is a healthy choice or may improve physical health
Code marketing strategy	
Corporate social responsibility or philanthropy	Statement of ethical or sustainable standpoint or initiative or charitable work undertaken by brand
Celebrities	People with an entertainment or media profile, excluding athletes
Sportspeople	Any person showing their athletic ability and/or sporting achievements, including extreme and motorsports
Children's characters	Third-party cartoons or characters, including characters from films, books, TV, and the internet
Branded characters	Any characters developed by the brand
Special price promotions	Limited time offers, discount menus, 2 for 1 deals, or other reduced price advertisements
Vouchers	"Offers" exclusively available to those who like the account, including print off and/or electronic codes
Competitions	Any contest involving participant entry, including minimal requirements, for example, liking a post
Engagement	Posts that prompt interaction/conversation
Sponsorships or partnerships	Any events the brand supports or brands/service partners, excluding charitable organizations (coded as corporate social responsibility, see previously)
Videos or Graphics Interchange Format (GIFs) or boomerang	Moving images
Links	Link to an external page or additional content
Branding elements	Logos, colors, fonts, trademarks, or slogans
Product imagery (unbranded)	Pictures of the products sold or their ingredients, with no labels or branding elements
Image with no clear marketing strategy	Image not associated with any other marketing category

When determining which codes to assign, post elements were examined systematically. First, the image was viewed, and then, if the primary marketing strategy could not be determined based on image alone, the original image caption was read (the first comment found under an Instagram post and made by the account owner). Any further comments by followers or the account owner were not incorporated.

Analysis

Interrater reliability was established by both coders independently coding the same random sample of 10 posts from each account (excluding Ben and Jerry's, n=140). A percentage agreement score was then calculated between the 2 coders. Percentage agreement of 80% or above was considered acceptable.

Descriptive statistics were generated in Excel (Microsoft) for each individual account for the average number of posts, average number of post likes and comments, and average number of views of videos (views are not published for image-based posts). Proxy interaction between account and consumer was determined by calculating the percentages of followers actively interacting with the account via liking or commenting on posts, or passively interacting by viewing videos. Instagram counts a view when a video is watched for at least 3 seconds. Viewing data were only available for videos posted after November 20, 2015, as this was when Instagram began releasing this information. All videos containing viewing data were included in this aspect of the analysis, not just those that were coded as video for their primary marketing strategy. The most common marketing strategies and the range of strategies used by each account were also determined.

Results

Brand Characteristics

Of the 15 Instagram accounts included in the study, there was one ice cream brand (Ben and Jerry's), one sweet biscuit (Oreo), and one sweet spread (Nutella), which were combined into a sugary grocery store foods category. There were also 2 soft drink brands (Coca-Cola and Pepsi), 3 energy drinks (Gatorade, Monster Energy, and Red Bull), and 7 fast food chains (Burger King, Dominos, KFC, McDonald's, Starbucks, Subway, and Taco Bell). The characteristics of each Instagram account, such as page likes and number of posts during the study period, are summarized in Table 2.

At the time of the study, Starbucks was the most popular account with 6.6 million followers, and Subway was the least popular in our sample with 109K followers. There was a large variation in the number of posts by each account, with some accounts (Dominos, Monster Energy, and Red Bull) posting multiple times a day, and other accounts (Oreo, Subway, and Pepsi) only posting every 3 to 4 days.

Table 2. Characteristics of top food and beverage Instagram accounts.

Rank	Account name and Instagram handle	Followers, n (Nov 2015)	Posts during study period, n	Average likes per post	Followers who liked post, %	Average comments per post	Followers who commented, %	Videos with views data, n	Average views per video	Followers who viewed video, %
1	Starbucks Coffee	6.6M	312	204,833	3.10	1318	0.02	10	518,903	7.86
2	Red Bull	3.8M	419	74,261	1.95	803	0.02	38	380,592	10.02
3	Monster Energy	2.6M	685	36,888	1.42	178	0.01	61	101,293	3.90
4	OREO	1.1M	94	31,510	2.86	628	0.06	20	110,828	10.08
5	McDonald's	1M	128	21,978	2.20	806	0.08	11	104,730	10.47
6	Coca-Cola	920K	117	19,333	2.10	317	0.03	24	53,150	5.78
7	Gatorade	678K	226	12,489	1.82	82	0.01	8	34,306	5.06
8	Taco Bell	671K	183	18,036	2.69	810	0.12	7	71,454	10.65
9	KFC	632K	160	7,697	1.22	235	0.04	8	30,884	4.89
10	Ben and Jerry's	588K	295	17,470	2.97	609	0.10	6	50,180	8.53
11	Burger King	530K	186	9,666	1.82	411	0.08	2	30,628	5.78
12	Domino's Pizza	431K	407	5,722	1.33	159	0.04	2	32,078	7.44
13	Pepsi	320K	136	4,545	1.42	610	0.19	9	15,523	4.85
14	Nutella	301K	226	14,260	4.74	373	0.12	16	66,150	21.98
15	Subway	109K	98	2,025	1.86	113	0.10	3	12,313	11.30
Average of all brands		1,352,000	245	32,048	2.37	497	0.04	15	107,534	7.95

Table 3. Marketing strategies used by brands on Instagram.

Instagram account	Top 3 most used marketing strategies (posts)			Marketing strategies (N=15), n	Informational posts, n (%)	Original posts, n (%)	Health claims, n
	1	2	3				
Sugary grocery store foods							
Ben and Jerry's (n=295)	Branding (123)	Links (66)	Corporate/social responsibility (38)	10	251 (85.1)	255 (86.4)	1
Nutella (n=226)	Branding (152)	Product image (65)	Competition and engagement (3)	6	219 (96.9)	218 (96.5)	0
Oreo (n=94)	Product image (48)	Branding (23)	Videos (11)	8	81 (86)	91 (97)	0
Soft drinks							
Coca-Cola (n=117)	Branding (51)	Videos (34)	Children's characters (17)	8	70 (59.8)	92 (78.6)	1
Pepsi (n=136)	Branding (88)	Competition (19)	Videos (19)	8	84 (61.8)	124 (91.2)	0
Energy drinks							
Gatorade (n=226)	Sportspeople (124)	Branding (77)	Image with no clear strategy (12)	7	96 (42.5)	226 (100)	9
Monster Energy (n=685)	Sportspeople (552)	Branded character (59)	Branding (28)	9	20 (2.9)	676 (98.7)	3
Red Bull (n=419)	Sportsperson (310)	Branding (31)	Videos (25)	11	15 (3.6)	416 (99.3)	0
Fast food chains							
Burger King (n=186)	Branding (99)	Product image (59)	Partnerships (17)	8	169 (90.9)	165 (88.7)	0
Dominos (n=407)	Product image (232)	Branding (101)	Image with no clear strategy (33)	11	302 (74.2)	407 (100)	4
KFC (n=160)	Branding (45)	Branded characters (26)	Videos (22)	10	91 (56.8)	160 (100)	0
McDonald's (n=128)	Branding (56)	Product image (35)	Videos (22)	7	108 (84.4)	119 (93.0)	0
Starbucks (n=312)	Branding (175)	Product image (67)	Links (26)	10	242 (77.6)	228 (73.1)	2
Subway (n=98)	Branding (50)	Product image (29)	Special price promotion (6)	6	70 (71)	97 (99)	2
Taco Bell (n=183)	Branding (88)	Product image (45)	Special price promotion (11)	11	140 (76.5)	182 (99.5)	1

Only a small percentage of followers actively interacted with the branded accounts via liking (1.22%-4.74%) or commenting (<1%) on their posts. Since November 2015, viewing data is publically available for video posts only, and this gives some indication of the level of passive interaction between consumers and branded accounts. Within each Instagram account, a greater percentage of followers passively interacted with account videos (up to 21.98%), compared with actively liking or commenting on posts. However, some accounts only had a very small number of videos with views data within the study period, eg, Burger King and Dominos with 2 video posts each. From the descriptive data in Table 2, there also appears to be little relationship between frequency of posts and level of passive or active interaction between account and account followers.

Marketing Strategies

Interrater reliability for the coding of Instagram posts was acceptable, with percentage agreement scores for accounts coded by individual coders ranging between 80% and 100% for all categories. A summary of marketing strategies and other characteristics of images posted are summarized in Table 3, and the full results of the marketing categorization for all images are in Multimedia Appendix 1.

Each brand used multiple primary marketing strategies in their Instagram accounts, ranging from 6 to 11 different strategies each. Some brands, particularly energy drinks (Gatorade, Monster Energy, and Red Bull), predominately favored certain strategies, with up to 81% of their posts featuring sportspeople. There was also a high level of branding contained in posts, with branding featuring as 1 of the top 3 marketing strategies for each account in this study, although this category was the second last in the coding hierarchy. Use of children's characters was 1 of the top 3 marketing strategies of just one brand, Coca-Cola.

The level of informational content provided on Instagram was highly variable by brand, ranging from 2.9% to 96.9% of images for each account. Energy drink brands had the lowest number of posts containing product or brand information, and sugary grocery food, particularly Nutella, products had the highest.

The majority of Instagram posts featured original images taken by the brand or a contracted photographer (73.1%-100% of images of each account). There were few identified "regrammed" images, where a brand reposts an image taken by a brand consumer and credits (or "tags") the image back to them.

There were very few health claim–related images posted on Instagram. Exceptions to this were Subway and Dominos occasionally highlighting the salad components of their product; Starbucks focusing on the fruit content of their smoothies; and Gatorade emphasizing the thirst quenching, electrolyte, and rehydration qualities of their energy drinks. "Healthy alternative" menu items featured occasionally but not prominently in any of the Instagram accounts studied.

Case Study Examples

Three case studies are presented for the accounts with the most followers (Starbucks), the most posts (Monster Energy), and the greatest amount of passive and active interaction (Nutella).

Starbucks contained the highest percentage of regrammed images and images of everyday people consuming their products. In comparison with other accounts, Starbucks was relatively informational about the brand and the products they sell. Figure 1 is a typical Starbucks post and contains a relatable consumer holding an energy-dense product with the highly recognizable Starbucks cup.

Monster Energy posted multiple times a day during the study period. Like other energy drinks, this account predominately featured posts of sportspeople and athletic endeavors (Figure 2). This figure shows a typical post by Monster Energy and emphasizes high energy, fiery, and adventurous associations with the brand, which are appealing to the target audience of male millennials [29], but does not feature the beverage itself. Monster Energy had the lowest amount of informational posts (2.9%). However, the Monster branding "M" or colors frequently featured.

Nutella was the brand with the greatest amount of active interaction through followers liking or commenting on their images or viewing their videos. Nutella (Figure 3) also had the greatest percentage of posts that contained consumer information, such as what the product looks like or how to use it (96.9% of posts). Almost all their posts featured different ways to prepare and eat Nutella, as seen in this example image, demonstrating its perceived versatility and the many ways the product can be incorporated into a consumer's everyday diet. The terms breakfast and start the day were common in Nutella posts, as seen in this caption, as well as incorporation of fruit.

Figure 1. Starbucks example image.

249,716 likes 76w

starbucks Dressing like fall, celebrating like summer. #WinWin #IcedSaltedCaramelMocha

view all 1,949 comments

kiara_diaz12 My sister likes starbucks

_carmennnn__ looks like something I don't have □□ @anthony.lara710

jenniferhrod @keepersb yes yes it does

_kaloodi_sw يعوه

romymulderr @rogierberendsen

cizemlermi @miraylermi

lermirayy Icilir☺ @cizemlermi

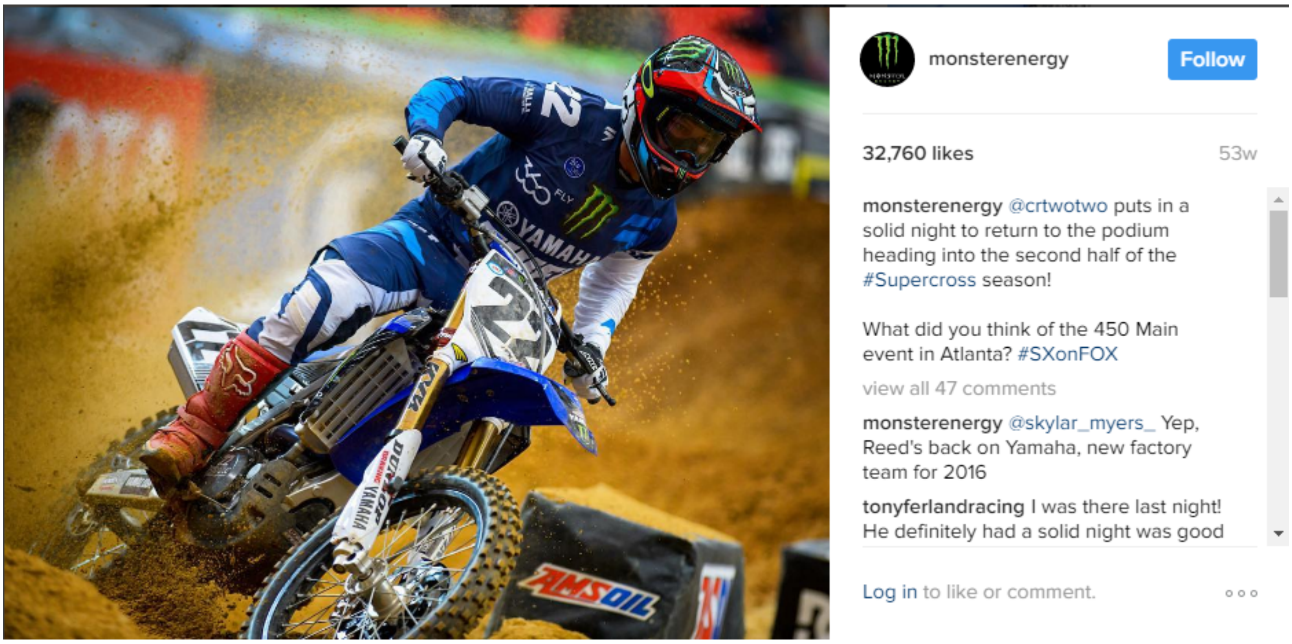
daiisyyy__ @flower_in_the_sun you got me ? Lmao

icantremembermypassworddamnit @m.kaylasaxton buy this for me

celescortez @emanuel_miranda94 *-*

Log in to like or comment.

Figure 2. Monster Energy example image.



monsterenergy Follow

32,760 likes 53w

monsterenergy @crtwotwo puts in a solid night to return to the podium heading into the second half of the #Supercross season!

What did you think of the 450 Main event in Atlanta? #SXonFOX

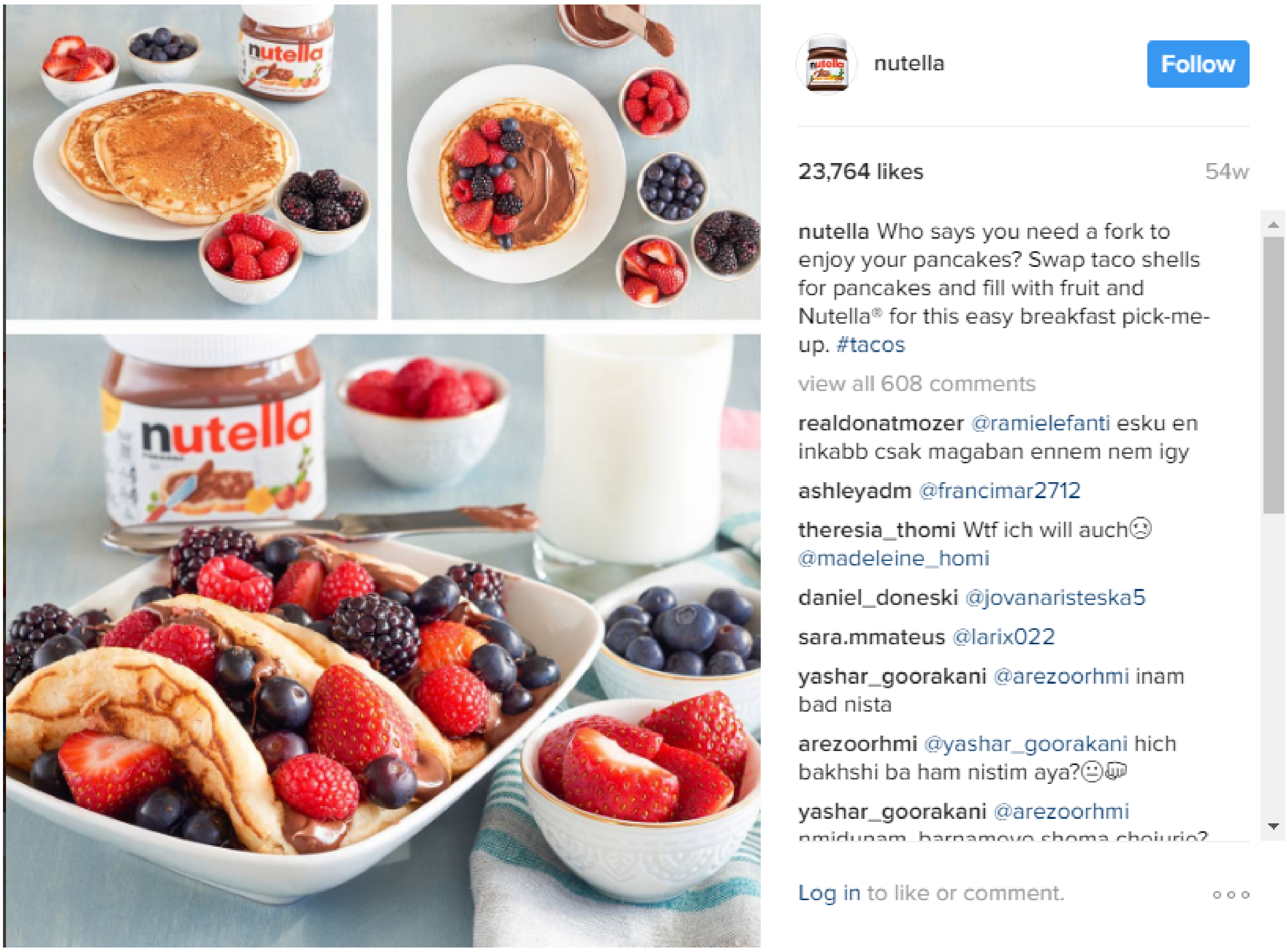
view all 47 comments

monsterenergy @skylar_myers_ Yep, Reed's back on Yamaha, new factory team for 2016

tonyferlandracing I was there last night! He definitely had a solid night was good

Log in to like or comment. ○○○

Figure 3. Nutella example image.



nutella Follow

23,764 likes 54w

nutella Who says you need a fork to enjoy your pancakes? Swap taco shells for pancakes and fill with fruit and Nutella® for this easy breakfast pick-me-up. #tacos

view all 608 comments

realdonatmozer @ramielefanti esku en inkabb csak magaban ennem nem igy

ashleyadm @francimar2712

theresia_thomi Wtf ich will auch☹ @madeleine_homi

daniel_doneski @jovanaristeska5

sara.mmateus @larix022

yashar_goorakani @arezoorhmi inam bad nista

arezoorhmi @yashar_goorakani hich bakhshi ba ham nistim aya?😐

yashar_goorakani @arezoorhmi

Log in to like or comment. ○○○

Discussion

Principal Findings

This study demonstrates the widespread use of Instagram to market ENDP food and beverages to consumers. The percentage of posts containing any informational content ranged from 2.9% to 96.9%, depending on the brand. Our definition was generous and included any information about the product, what it looks or tastes like, its ingredients, or how it could be purchased. We did not assess quality of the information nor did we find any evidence that brands were sharing important nutritional information about caloric energy, sugar, salt, and fat content, or serving size. In addition, this information was accompanied by an image or video that also included persuasive marketing strategies. Energy drinks, in particular, had very low numbers of informational posts and instead focused on the symbolic uses of their products via extreme sports, rather than including any information on what their product actually is, looks, or tastes like. These results support the conclusion that advertising places less emphasis on communicating specific product information and more on communicating the social and symbolic uses of products [14] and building relationships with consumers [24].

Similar to other marketing platforms, branded characters, symbols, and colors were used throughout the Instagram posts for all brands included in this study. Previous research has found that familiar and/or character branding has a powerful influence on children's preferences, choices, and intake of foods, especially unhealthy foods [30]. Celebrity endorsement is a common and longstanding strategy used in many advertising platforms [11]. This study found that athletes and celebrities were commonly incorporated into Instagram images, particularly by energy drink brands. Celebrities also have their own Instagram accounts, which are used to promote particular brands through personal testimonies [31], further expanding the reach and persuasiveness of brand advertising, although these pages were not analyzed in this study.

The relatively small percentage of followers who liked or commented on posts demonstrated potentially low levels of active interaction initiated by consumers with brands on Instagram. This aligns with previous research demonstrating that many on social media are passive users who view content but do not actively engage [28]. Yet, despite this, the frequency of posting on Instagram as well as the ability to sponsor, promote, or search for posts and hashtags [32] suggests that consumers may be exposed to a large amount of EDNP food images, especially if they follow multiple big brands and login multiple times a day [33].

This study found a low prominence of healthy products among the most popular brands and few health-related claims on Instagram. Although brands are not directly advertising their products as healthy, they do imply that their products are part of a balanced diet. For example, Nutella commonly uses hashtags like breakfast, or statements like Nutella for breakfast never goes out style. The brands under study also did not promote their healthy options on Instagram. Companies, in particular fast food restaurants, have previously stated that their healthy options are priority products [34]; however, the results of this study do not support this claim. Use of social media also allows companies to bypass promises regarding advertising to children and young people. For example, a spokesperson for McDonald's has previously commented that the brand takes its responsibility as an advertiser seriously and that they only promote healthier options when communicating to children [35]. Although this may be the case for television advertising during children's programs, these promises are not being adhered to on Instagram, a finding similar to previous research relating to alcohol marketing to youth [36]. Although Instagram's terms of use require users to be aged older than 13 years, and therefore brands are operating within the terms of these responsibility statements, statistics from the United Kingdom show that in 2016, approximately 50% of children aged 8 to 15 years use this social media platform [37].

The unique, individual, and highly curated nature of each Instagram account included in this study demonstrates that different audiences are being targeted through these different accounts. Market research has shown that consumers, particularly those of the digital age, are more receptive to advertising that is tailored to their needs and interests and inserted into their personalized media experiences [10]. This exemplifies the extraordinary potential of Instagram and other visually based social media platforms to effectively communicate with potential consumers and encourage regular purchase and consumption of unhealthy foods and beverages. The everyday nature of these posts and constant exposure to these images have the potential to condition consumers to disregard the sometimes or special occasion recommended use of these products and incorporate them as a normal or everyday part of their diet. There are also examples of chains developing new EDNP products based primarily on their photogenicity and ability to trend on Instagram. One example is the Starbucks Unicorn Frappuccino, with a Starbucks spokesperson stating that "the look of the beverage was an important part of its creation, our inspiration came from the fun, spirited and colorful unicorn theme food and drinks…trending on social media" [38].

Study Limitations

This study has some limitations. First, it was not possible to determine the number of views for each post, which would be valuable information to determine the reach of ENDP food advertising on Instagram. This information is only available to the account owners. In addition, percentage of followers was used in this study as a proxy measure for interaction between branded accounts and the public. However, because of Instagram's undisclosed algorithm for determining which images appear in what order on a user's newsfeed, individuals may not see every post from a branded account even if they follow them. Overall, these presented results are likely to be an underestimate of true exposure to these images, as nonfollowers are also able to view these posts through sharing, searching, hashtags, or the more recent Instagram explore feature. Only one primary marketing strategy was assigned to each Instagram post, as coding all applicable strategies to each image would have overcomplicated and reduced the meaningfulness of the results. Therefore, this study presents an underestimate of the use of some marketing strategies, particularly those toward the bottom of this hierarchy, and prevalence of the use of each strategy cannot be accurately reported. Moreover, this study only analyzed images posted on each brand's account; it did not incorporate paid advertisement or promoted posts, which appear on users' newsfeed whether or not they follow that brand. Finally, any assessment of the followers of each Instagram account was not incorporated into this study and is instead a topic for future research in this field.

Conclusions

Instagram is a social media platform where the promotion of EDNP foods and beverages is flourishing. This study has demonstrated a high frequency of advertising by top food and beverage brands on Instagram through targeted and curated posts using a range of marketing strategies. Brands are using social media platforms such as Instagram to advertise to a growing number of consumers, with little to no restrictions. Public health bodies need to act to develop compelling social counter-marketing campaigns as a way of combating this advertising [39]. This study also contributes to the growing evidence of the urgent need to establish policy action regarding ENDP food and beverage advertising [40] that is reflective of the current media environment.

Acknowledgments

The authors acknowledge Ms Tala Barakat for assistance with establishing the image coding process. This work was supported by a Preventive Health Research Grant (grant number 64FRE2011).

Authors' Contributions

AJV designed the study, analyzed and interpreted the data, drafted the manuscript, and approved the final paper. BK assisted with interpretation of the data, critically revised the manuscript drafts, and approved the final paper. LZ and ZW acquired the data, critically revised the manuscript drafts, and approved the final paper. SY assisted with data coding, critically revised the manuscript drafts, and approved the final paper. BF conceived the study, assisted with analysis and interpretation of the data, critically revised the manuscript drafts, and approved the final paper.

Conflicts of Interest

BF has the following potential conflicts of interest to disclose: consultancy (money paid to author) from WHO, Sax Institute, Cancer Council NSW, NSW Department of Health, Cancer Institute; consultancy (money paid to author's institute) from Australian Department of Health, NSW Department of Health, Cancer Institute. Grants paid to author's institution from NHMRC, ANPHA; and travel, accommodation, meeting expenses paid to author's institution from UICC and National Taiwan University.

Multimedia Appendix 1

All coded data for marketing strategies used by brands on Instagram.

[PDF File (Adobe PDF File), 145KB - publichealth_v4i2e54_app1.pdf]

References

1. World Health Organization. 2016. Obesity and overweight factsheet URL: http://www.who.int/en/news-room/fact-sheets/detail/obesity-and-overweight [accessed 2018-05-10] [WebCite Cache ID 6zJ0grRKE]
2. Swinburn BA, Sacks G, Hall KD, McPherson K, Finegood DT, Moodie ML, et al. The global obesity pandemic: shaped by global drivers and local environments. Lancet 2011 Aug 27;378(9793):804-814. [doi: 10.1016/S0140-6736(11)60813-1] [Medline: 21872749]
3. Ng M, Fleming T, Robinson M, Thomson B, Graetz N, Margono C, et al. Global, regional, and national prevalence of overweight and obesity in children and adults during 1980-2013: a systematic analysis for the Global Burden of Disease Study 2013. Lancet 2014 Aug 30;384(9945):766-781. [doi: 10.1016/S0140-6736(14)60460-8] [Medline: 24880830]
4. Chandon P, Wansink B. Does food marketing need to make us fat? a review and solutions. Nutr Rev 2012 Oct;70(10):571-593 [FREE Full text] [doi: 10.1111/j.1753-4887.2012.00518.x] [Medline: 23035805]
5. Dixon H, Scully M, Niven P, Kelly B, Chapman K, Donovan R, et al. Effects of nutrient content claims, sports celebrity endorsements and premium offers on pre-adolescent children's food preferences: experimental research. Pediatr Obes 2014 Apr;9(2):e47-e57. [doi: 10.1111/j.2047-6310.2013.00169.x] [Medline: 23630014]
6. Boyland EJ, Nolan S, Kelly B, Tudur-Smith C, Jones A, Halford JC, et al. Advertising as a cue to consume: a systematic review and meta-analysis of the effects of acute exposure to unhealthy food and nonalcoholic beverage advertising on intake in children and adults. Am J Clin Nutr 2016 Feb;103(2):519-533 [FREE Full text] [doi: 10.3945/ajcn.115.120022] [Medline: 26791177]
7. Clark CR, Doraszelski U, Draganska M. The effect of advertising on brand awareness and perceived quality: an empirical investigation using panel data. Quant Mark Econ 2009 May 8;7(2):207-236. [doi: 10.1007/s11129-009-9066-z]
8. Wilcox B, Kunkel D, Cantor J, Dowrick P, Linn S, Palmer E. American Psychological Association. 2004. Report of the APA task force on advertising and children URL: https://www.apa.org/pi/families/resources/advertising-children.pdf [accessed 2018-05-10] [WebCite Cache ID 6zJ0uorAb]
9. Freeman B. New media and tobacco control. Tob Control 2012 Mar;21(2):139-144. [doi: 10.1136/tobaccocontrol-2011-050193] [Medline: 22345236]
10. Montgomery KC, Chester J. Interactive food and beverage marketing: targeting adolescents in the digital age. J Adolesc Health 2009 Sep;45(3 Suppl):S18-S29. [doi: 10.1016/j.jadohealth.2009.04.006] [Medline: 19699433]

11. Keel A, Nataraajan R. Celebrity endorsements and beyond: new avenues for celebrity branding. Psychol Mark 2012 Aug 28;29(9):690-703. [doi: 10.1002/mar.20555]
12. Cairns G, Angus K, Hastings G. The extent, nature and effects of food promotion to children: a review of the evidence to December 2008. Geneva, Switzerland: World Health Organization; 2009.
13. Chester J, Montgomery K. Digital ads. 2007. Interactive food and beverage marketing: targeting children and youth in the digital age URL: http://digitalads.org/documents/digiMarketingFull.pdf [accessed 2018-05-10] [WebCite Cache ID 6zJ1FpDeV]
14. Jolly R. Canberra: Parliament of Australia; 2011. Marketing obesity? junk food, advertising and kids URL: https://www.aph.gov.au/About_Parliament/Parliamentary_Departments/Parliamentary_Library/pubs/rp/rp1011/11rp09 [accessed 2018-05-10] [WebCite Cache ID 6zJ1NAHEb]
15. Worthan J. The New York Times. 2012. Facebook to buy photo-sharing service Instagram for $1 Billion URL: https://bits.blogs.nytimes.com/2012/04/09/facebook-acquires-photo-sharing-service-instagram/ [accessed 2018-12-07] [WebCite Cache ID 1512607933446000]
16. Instagram. 2017. 700 Million URL: http://blog.instagram.com/post/160011713372/170426-700million [accessed 2018-05-25] [WebCite Cache ID 6vWKYf5nT]
17. Perkins E. PACE. 2016. Social media network demographics: what brands need to know URL: https://www.paceco.com/insights/social-media/popular-social-media-demographics/ [accessed 2017-12-07] [WebCite Cache ID 6vWKhJZzB]
18. Greenwood S, Perrin A, Duggan M. Pew Research Center. 2016. Social media update 2016 URL: http://www.pewinternet.org/2016/11/11/social-media-update-2016/ [accessed 2018-02-22] [WebCite Cache ID 6vWL4yf9F]
19. Krogstad JM. Pew Research Center. 2015. Social media preferences vary by race and ethnicity URL: http://www.pewresearch.org/fact-tank/2015/02/03/social-media-preferences-vary-by-race-and-ethnicity [accessed 2017-12-07] [WebCite Cache ID 6vWL7S6cJ]
20. Sensis. 2015. Sensis social media report may 2015: how Australian people and businesses are using social media URL: https://www.sensis.com.au/asset/PDFdirectory/Sensis_Social_Media_Report_2015.pdf [accessed 2018-05-10] [WebCite Cache ID 6zJ2il4pt]
21. Castillo M. CNBC. 2017. Instagram tries to steal ad dollars from snap by proving it drives sales URL: http://www.cnbc.com/2017/03/21/instagram-proving-advertising-roi.html [accessed 2018-03-29] [WebCite Cache ID 6vWLN7p0m]
22. Freeman B, Kelly B, Baur L, Chapman K, Chapman S, Gill T, et al. Digital junk: food and beverage marketing on Facebook. Am J Public Health 2014 Dec;104(12):e56-e64. [doi: 10.2105/AJPH.2014.302167] [Medline: 25322294]
23. Coary S, Poor M. How consumer-generated images shape important consumption outcomes in the food domain. JCM 2016;33(1):1-8. [doi: 10.1108/JCM-02-2015-133]
24. Ginsberg K. Elon University. 2015. Instabranding: shaping the personalities of the top food brands on instagram URL: https://www.elon.edu/u/academics/communications/journal/wp-content/uploads/sites/153/2017/06/08GinsbergEJSpring15.pdf [accessed 2018-05-10] [WebCite Cache ID 6zJ2uoa3n]
25. Cohen J. Forbes. 2015. Eight reasons all brands should be on instagram URL: http://www.forbes.com/sites/jennifercohen/2015/05/05/8-reasons-all-brands-should-be-on-instagram/#51bdbddc447b [accessed 2018-02-22] [WebCite Cache ID 6vWLOO3y4]
26. Mander J. Global Web Index. 2015. Half of instagrammers follow brands URL: https://blog.globalwebindex.net/chart-of-the-day/half-of-instagrammers-follow-brands/ [accessed 2017-12-07] [WebCite Cache ID 6vWLQcNGL]
27. Cherian R, Westbrook M, Ramo D, Sarkar U. Representations of codeine misuse on instagram: content analysis. JMIR Public Health Surveill 2018 Mar 20;4(1):e22. [doi: 10.2196/publichealth.8144] [Medline: 29559422]
28. Kite J, Foley BC, Grunseit AC, Freeman B. Please like me: facebook and public health communication. PLoS One 2016;11(9):e0162765 [FREE Full text] [doi: 10.1371/journal.pone.0162765] [Medline: 27632172]
29. Heckman M, Sherry K, Mejia D, Gonzalez E. Energy drinks: an assessment of their market size, consumer demographics, ingredient profile, functionality, and regulations in the United States. Compr Rev Food Sci Food Saf 2010;9(3):303-317 [FREE Full text] [doi: 10.1111/j.1541-4337.2010.00111.x]
30. Kraak VI, Story M. Influence of food companies' brand mascots and entertainment companies' cartoon media characters on children's diet and health: a systematic review and research needs. Obes Rev 2015 Feb;16(2):107-126 [FREE Full text] [doi: 10.1111/obr.12237] [Medline: 25516352]
31. Maheshwari S. The New York Times. 2016. Endorsed by a Kardashian, but is it love or just an ad? URL: https://www.nytimes.com/2016/08/30/business/media/instagram-ads-marketing-kardashian.html?_r=0 [accessed 2018-04-05] [WebCite Cache ID 6vWLaUkYw]
32. Guidry JD, Messner M, Jin Y, Medina-Messner V. From #mcdonaldsfail to #dominossucks. Corp Comm: An Int Jnl 2015 Aug 03;20(3):344-359. [doi: 10.1108/CCIJ-04-2014-0027]
33. Byrne M. Opticalcortex. 2014. Instagram statistics optical cortex URL: https://opticalcortex.com/instagram-statistics/ [accessed 2017-12-07] [WebCite Cache ID 6vWLcm2ZQ]
34. Kitney D. The Australian. 2016. Healthy options take priority among major US food producers URL: http://www.theaustralian.com.au/business/in-depth/global-food-forum/healthy-options-take-priority-among-major-us-food-producers/news-story/953720ed5eecb0f686ca1a41d455b40d [accessed 2018-05-08] [WebCite Cache ID 6vWLoiBM7]

35. Han E. The Sydney Morning Herald. 2017. Food industry's voluntary codes failing to reduce junk food ads viewed by children URL: http://www.smh.com.au/business/consumer-affairs/food-industrys-voluntary-codes-failing-to-reduce-junk-food-ads-viewed-by-children-20170214-gucm4c.html [accessed 2017-12-07] [WebCite Cache ID 6vWLwCjyw]
36. Barry AE, Bates AM, Olusanya O, Vinal CE, Martin E, Peoples JE, et al. Alcohol marketing on twitter and instagram: evidence of directly advertising to youth/adolescents. Alcohol Alcohol 2016 Jul;51(4):487-492. [doi: 10.1093/alcalc/agv128] [Medline: 26597794]
37. Ofcom. 2016. Children and parents: media use and attitudes report 2016 URL: https://www.ofcom.org.uk/research-and-data/media-literacy-research/childrens/children-parents-nov16 [accessed 2018-05-10] [WebCite Cache ID 6zJ3FHttE]
38. Kulp P. Mashable. 2017. Starbucks' Unicorn Frappuccino is the latest example of how social media makes chains get weird URL: http://mashable.com/2017/04/23/starbucks-unicorn-frappuccino-chains-instagram-bait/ [accessed 2017-12-07] [WebCite Cache ID 6vWM5N7Cs]
39. Bellew W, Bauman A, Freeman B, Kite J. Social countermarketing: brave new world, brave new map. J Soc Mark 2017;7(2):205-222 [FREE Full text] [doi: 10.1108/JSOCM-09-2016-0052]
40. Nestle M. Food marketing and childhood obesity--a matter of policy. N Engl J Med 2006 Jun 15;354(24):2527-2529. [doi: 10.1056/NEJMp068014] [Medline: 16775233]

Abbreviations

EDNP: energy-dense, nutrient-poor

Edited by G Eysenbach; submitted 06.12.17; peer-reviewed by A Cyr, A McCosker; comments to author 17.03.18; revised version received 28.03.18; accepted 29.03.18; published 05.06.18.

Please cite as:
Vassallo AJ, Kelly B, Zhang L, Wang Z, Young S, Freeman B
Junk Food Marketing on Instagram: Content Analysis
JMIR Public Health Surveill 2018;4(2):e54
URL: http://publichealth.jmir.org/2018/2/e54/
doi:10.2196/publichealth.9594
PMID:

Original Paper

#Healthy Selfies: Exploration of Health Topics on Instagram

Sachin Muralidhara[1], MS; Michael J Paul[2], PhD

[1]Department of Computer Science, University of Colorado Boulder, Boulder, CO, United States

[2]Department of Information Science, University of Colorado Boulder, Boulder, CO, United States

Corresponding Author:
Michael J Paul, PhD
Department of Information Science
University of Colorado Boulder
315 UCB
Boulder, CO, 80309
United States
Phone: 1 303 735 7581
Email: mpaul@colorado.edu

Abstract

Background: Social media provides a complementary source of information for public health surveillance. The dominate data source for this type of monitoring is the microblogging platform Twitter, which is convenient due to the free availability of public data. Less is known about the utility of other social media platforms, despite their popularity.

Objective: This work aims to characterize the health topics that are prominently discussed in the image-sharing platform Instagram, as a step toward understanding how this data might be used for public health research.

Methods: The study uses a topic modeling approach to discover topics in a dataset of 96,426 Instagram posts containing hashtags related to health. We use a polylingual topic model, initially developed for datasets in different natural languages, to model different modalities of data: hashtags, caption words, and image tags automatically extracted using a computer vision tool.

Results: We identified 47 health-related topics in the data (kappa=.77), covering ten broad categories: acute illness, alternative medicine, chronic illness and pain, diet, exercise, health care & medicine, mental health, musculoskeletal health and dermatology, sleep, and substance use. The most prevalent topics were related to diet (8,293/96,426; 8.6% of posts) and exercise (7,328/96,426; 7.6% of posts).

Conclusions: A large and diverse set of health topics are discussed in Instagram. The extracted image tags were generally too coarse and noisy to be used for identifying posts but were in some cases accurate for identifying images relevant to studying diet and substance use. Instagram shows potential as a source of public health information, though limitations in data collection and metadata availability may limit its use in comparison to platforms like Twitter.

(JMIR Public Health Surveill 2018;4(2):e10150) doi:10.2196/10150

KEYWORDS

social media; Instagram; image sharing; topic modeling; computer vision; public health

Introduction

Background

Social media can provide a vast source of insight into a wide variety of applications in public health monitoring and surveillance [1]. The bulk of social media-based health monitoring has relied on Twitter, a microblogging platform with over 300 million active users worldwide [2]. A wide variety of health topics are openly discussed on Twitter [3], providing researchers with a rich source of data for monitoring the spread of disease [4,5], dietary patterns [6,7], drug abuse [8,9], foodborne illness [10,11], and depression [12,13], among many other applications.

While Twitter has strengths as a data source, its dominance in research relative to other public platforms has been explained as a matter of convenience: Twitter provides free APIs to obtain large volumes of random or targeted samples of data [1]. However, microblogs are only one type of social media. Other social media platforms contain different types of data and are used in different ways, like sharing visual media. Image-sharing platforms, such as Flickr, Tumblr, Pinterest, and Instagram, are very popular; for example, Instagram, the most popular image-sharing platform, is more than twice the size of Twitter, with over 700 million active users [14]. Despite their popularity, relatively few public health studies have used these types of platforms as a data source [1].

Most prior health research using image-sharing platforms has focused on lifestyle issues, such as diet and substance use. Mejova et al [15] analyzed posts of food on Instagram, focusing on the relationship between food consumption and obesity. De Choudhury et al [16] also examined food consumption on Instagram, focusing specifically on dietary patterns in locations classified as “food deserts.” Yom-Tov et al [17] and Pless et al [18] examined imagery associated with eating disorders on Flickr and Tumblr, respectively. A few studies have looked at substance use on Instagram, including electronic cigarettes [19], marijuana [20], and opioids [21], as well as the marketing of substances on Instagram [22,23]. Garimella et al [24] looked more broadly at lifestyle choices in Instagram, including diet, physical activity, and drinking.

Many of these studies focused on the text features (eg, hashtags and captions) of the image posts, and some conducted a manual content analysis of the images themselves. Two of the studies cited above-used computer vision—a type of artificial intelligence that can automatically analyze the content of images—to perform automatic identification of certain types of images. Pless et al [18] built image classification models to identify images promoting anorexia, since such content may not be tagged with informative text captions. Garimella et al [24] attempted to estimate county-level health statistics from social media content and experimented with both text features and automatically-extracted image tags, finding that both types of information could be correlated with external health metrics.

This study seeks to characterize the health content shared on Instagram, the most popular image-sharing site, toward the goal of identifying potential areas of research that may benefit from this type of data source. In particular, we consider the following research questions: (1) what health topics are prominently shared in Instagram, and (2) what are the characteristics of those topics, specifically the types of images associated with the topics? This study is related to exploratory topic analyses of other platforms for health research [3,25], with an additional contribution of characterizing the features of images in addition to text. The dataset is made available as a resource to the public health informatics community.

Methods

Ethics Statement

This study was reviewed by the University of Colorado Boulder Institutional Review Board, which determined that it does not constitute human subjects research. However, given that publicly posted images may still be considered sensitive material by the users, we took steps to preserve privacy. We did not download any images as part of our study. Instead, we collected the URL pointers to images hosted on Instagram, which were processed by an external computer vision application programming interface (API). Our data collection only contains the abstract features extracted from the images.

Data Collection and Preparation

While Instagram provides APIs for specific applications, Instagram does not provide an API to collect public data [26], in contrast to Twitter which provides widely-used streaming APIs [27]. We instead built a “crawler” that queries Instagram’s search engine, which returns the nine most recent posts matching a specified hashtag. The crawler accesses the webpage of the hashtag search engine, analogous to how a person would access the search engine in their browser. The page is downloaded, and the HTML is parsed to extract information such as the set of tags and the caption of the image.

We iteratively queried the search engine for 269 general health-related keywords that were used in previous work to obtain a general collection of health-related tweets [3]. The keywords were obtained from dictionaries of terms related to diseases, symptoms, and treatments, in addition to general words like “sick” and “health” that were added manually. The original keyword set contained over 20,000 terms, which were reduced to 269 words that were most common in Twitter, to conform to API limits on how many keywords can be searched. While developed for Twitter, we use the same list for Instagram here, as the list contains a broad set of terms that have previously been shown to be useful for collecting health-related social media posts.

Instagram only allows searching for hashtags rather than free text, so we treated each keyword as a hashtag (eg, “#flu” instead of “flu”). We repeated these 269 queries continuously from September 29, 2016, through October 25, 2016, attempting to simulate a “streaming” collection as with Twitter, and obtained 174,517 posts. We did not download the same post more than once as measured by a unique post identifier. However, if the same content was shared in multiple posts (eg, if multiple users shared the same image), these posts would be considered separate in the dataset.

Each post includes an image, a set of hashtags, and an optional free text caption. We used langid.py to identify and remove posts containing non-English captions [28]. We also removed hyperlinks, and nonalphanumeric characters. Stop words were removed using the natural language toolkit (NLTK) [29]. After filtering and processing, our dataset contained 96,426 documents posted by 77,327 users with an average of 1.25 posts per user.

We extracted “tag” features from each image using Microsoft’s Computer Vision API [30]. This service returns key phrase descriptors of images, such as “person” or “running”. The API was able to extract at least 1 tag for 79.24% (76,407/96,426) of images in our dataset. We refer to these tags as “image tags” to distinguish them from hashtags.

Once extracted, we treated the image tags as an additional type of text, along with captions and hashtags. In the final collection, there are 96,426 posts with a nonempty list of hashtags for an average of 15.2 tags per post. There were 95,208 posts (95,208/96,426; 98.74%) with nonempty captions for an average of 21.6-word tokens per caption. There were also 76,407 posts (76,407/96,426; 79.24%) with at least 1 image tag for an average of 3.7 tags per image. The dataset is shown in Multimedia Appendix 1.

Topic Modeling

We use probabilistic topic models [31] to characterize the major themes of health-related discussion in Instagram. Topic models are tools for clustering related words into themes or concepts called "topics" and for identifying the topic composition of documents. Topic models have been used in health research as a method of performing content analyses of large datasets [25,32-34].

A topic model is a statistical model with many latent variables and parameters that can be inferred by fitting the model to data. In this model, each "topic" has a probability distribution over words, estimated from data, and topics are usually represented by presenting the 10-20 most probable words in the topic. Additionally, each document has a probability distribution over topics, which can be used to characterize the topic composition of a document and to identify documents that describe particular topics.

Topic models take documents, represented as vectors of word counts, as input. The model parameters (ie, the distribution over topics θ_d in each document *d*, and the distribution over words φ_k for each topic *k*) are estimated to fit the observed data (ie, the word counts in each document). The estimated parameters are often interpretable to people, and the words associated with each topic cluster can be used to assign a concept to the topic [35].

Polylingual Topic Model

The polylingual topic model [36] is an extension of a traditional topic model that is applied to multiple languages. This model can be used for datasets in which documents have multiple versions in multiple languages. For example, translations of a document into other languages, or articles in different languages that are known to be about the same topic, like different versions of Wikipedia articles. In the polylingual topic model, the distribution over topics θ_d is shared across all versions of the document, while each topic has a different distribution over words specific to each language *l*, φ_{lk}.

In this work, we treat the different modalities of data—captions, hashtags, and image tag features—as different "languages" and apply the polylingual topic model to these 3 types of data. That is, each topic has a distribution over caption words, a distribution over hashtags, and a distribution over image tags. This will provide different views of each topic, allowing us to leverage multiple types of data and provide a complete understanding of the topics.

Model Estimation

We used the Polylingual Topic Model implementation from MALLET [37]. The hyperparameter for the topic distribution prior (ie, "alpha") was set to 1.0, and we used the default algorithm settings. The number of topics was set to 150. The model does not require each document to have a version in all 3 "languages," and if a document did not contain a caption or image tags, we still included the document but without those data types.

Because the topic model output in this study is interpreted qualitatively to be used in a content analysis, we also used qualitative judgment in performing model selection [1]. To avoid extensive model selection, we relied on default hyperparameters for the model. To choose the number of topics, we compared the output with 50, 100, and 150 topics. We selected 150 topics because this setting provided topics that were qualitatively more coherent.

Topic Identification

After running the topic model, we examined the 20 most probable words in each "language" of each topic. The two authors independently annotated each topic, labeling each topic with a phrase that describes the group of words or marking the topic with an "unknown" label if the words do not form a coherent theme. The annotators then discussed the independent labels with each other to determine if the 2 labels described the same concept (eg, the free text labels could be similar but different strings, such as "Running" and "Jogging"), and to decide on a final label.

When comparing whether the 2 annotators thought a topic was coherent, as opposed to the "unknown" label, the annotators agreed on 124/150 (82.7%) of the topics (Cohen kappa=.62). When comparing whether the 2 annotators thought a topic was related to health, the agreement was 136/150 (90.7%) with Cohen kappa=.77.

Additionally, we grouped the topics into coarse-grained categories, to make the results easier to summarize. One annotator created a grouping of the topics and then iterated with feedback from the other annotator. Categories are not mutually exclusive; topics could be assigned to more than 1 category.

Topic Analysis

The topic model gives the probability of each topic *k* in each document *d*, θ_{dk}. To get the overall prevalence of each topic, we calculate the average proportion, where *D* is the number of documents:

To estimate the prevalence of the coarser categories, we simply define each category's prevalence of the sum of its topic proportions, where C_j is the set of topics in category *j*:

To summarize the degree to which 2 topics have a tendency to occur together in documents, we calculate the Pearson correlation of the θ_{dk} values for all pairs of topics. Most of the topic pairs with high correlations were similar topics in the same category. To discover less obvious topic co-occurrences, we focus only on topic pairs from different categories.

Results

Of all the topics, 93/150 (62%) were identified as coherent, with 47/150 (31%) related to health and 46/150 (31%) not related to health. The 47 health topics were grouped into 10 high-level categories: acute illness, alternative medicine, chronic illness and pain, diet, exercise, health care & medicine, mental health, musculoskeletal health and dermatology, sleep, and substance use.

Examples of health topics are shown in Table 1, while the complete set of 47 health topics organized across the 10 categories is provided in Multimedia Appendix 2. Examples of nonhealth topics are shown in Table 2. As can happen with unsupervised topic models, many clusters have very similar and overlapping content, with similar or identical names given by the annotators.

Comparing the different modalities, hashtags tend to contain words specific to the topic (eg, "cancer," "diabetes," "allergies"), while caption words give indications of the context of the posts (eg, "feel," "love," "hope," "proud"). In some cases, the caption words include first-person (eg, "i've," "i'll") and informal (eg, "awesome," "lol") language that might be observed in personal conversations. Other topics include caption words consistent with advertising (eg, "product," "call," "email," "consultation").

Examining the image tags, there are often not many tags that are directly related to the health topics, with the exception of topics related to food and beverage, which usually had explicit image descriptors of the corresponding food. For example, the *Meat* topic contains the image tags "meat" and "barbeque," *Desserts* contains "dessert" and "chocolate," *Alcohol* contains "alcohol" and "beer," and *Caffeine* contains "coffee" and "coffee cup." A small number of other topics are also associated with image tags that are directly related to the topic: *Sleep* includes "pillow" and "bed," and *Dental health* includes "toothbrush" and "mouth."

Table 1. The top 10 words in each modality (hashtags, caption words, and image tags) for 6 example health topics.

Modality	Bodybuilding	Cancer	Caffeine	Desserts	Insomnia	Suicide & self-harm
Hashtags	bodybuilding	cancer	caffeine	food	insomnia	anxiety
	shredded	cure	coffee	cravings	bedtime	depressed
	muscle	chemo	coffeelover	foodporn	workout	sad
	gym	breastcancer	coffeeaddict	delicious	art	suicide
	abs	cancersucks	coffeetime	chocolate	selfie	suicidal
	physique	breastcancerawareness	coffeeholic	foodie	rest	depression
	veins	pink	hot	yummy	night	cutting
	gymlife	chemotherapy	drink	dessert	natural	sadness
	bodybuilder	fuckcancer	cafe	sweet	sleepy	broken
	gains	hope	coffeegram	yum	amazing	selfharm
Captions	bro	cancer	coffee	chocolate	sleep	don't
	man	breast	day	cream	night	feel
	nice	awareness	tea	pumpkin	bed	talk
	work	month	today	eat	hours	people
	likes	pink	love	good	time	hate
	gym	support	drink	made	back	i've
	muscle	women	cup	butter	asleep	stop
	hard	chemo	hot	ice	sleeping	anymore
	follow	fight	green	cake	make	fucking
	training	family	feeling	peanut	nights	cry
Image tags	man	group	beverage	dessert	indoor	close
	underpants	standing	food	chocolate	lying	dark
	sport	people	coffee	slice	bedclothes	woman
	indoor	beautiful	table	food	smiling	staring
	barbell	crowd	drink	piece	pillow	clouds
	cellphone	little	coffeecup	eaten	blanket	hand
	phone	girl	breakfast	cream	sofa	cloudy
	holding	wearing	pastry	plate	bedroom	road
	exercisedevice	white	dixiecup	fork	glasses	nightsky
	swimsuit	pink	doughnut	pastry	cloth	mountain

Table 2. Examples of topics that are not directly about health.

Modality	Inspiration	Poetry & quotes	Spirituality	Politics	Cats	Grunge/emo
Hashtags	inspiration	poetry	heal	trump	catsofinstagram	grunge
	selflove	quotes	healing	vaccines	cat	tumblr
	happiness	pain	energy	nature	cute	emo
	recovery	words	meditation	vegan	cats	alternative
	positivity	quote	love	wakeup	kitty	depressed
	positivevibes	writer	spiritual	hillary	kitten	depression
	loveyourself	love	soul	blacklivesmatter	pet	goth
	heal	writersofinstagram	mind	usa	meow	sad
	hope	hurt	spirit	clinton	fluffy	aesthetic
	positive	writing	light	organic	animal	punk
Captions	life	love	energy	people	feel	don’t
	things	heart	healing	world	hope	feel
	live	words	body	trump	poor	i’ll
	past	world	soul	media	baby	back
	true	soul	light	american	glad	i’ve
	grateful	make	life	america	aww	make
	time	life	heart	vote	cat	yeah
	living	mind	deep	country	hear	hope
	mind	hurt	space	drugs	rest	feeling
	people	give	love	government	sick	man
Image tags	indoor	wearing	posing	indoor	domesticcat	indoor
	posing	day	person	screen	mammal	close
	sky	mammal	fresh	display	laying	person
	rock	dark	forest	electronics	animal	young
	garden	shore	mountain	flat	sleeping	hair
	water	grass	mammal	suit	cat	blue
	mountain	open	silhouette	sign	white	glasses
	bushes	plaque	sunset	text	rodent	looking
	clouds	building	sign	newspaper	grey	messy
	can	abstract	distance	computer	gray	silhouette

While not explicitly about the health topic, some topics contain image tags that convey other characteristics of the imagery. For example, the image tags of the *Mental health* topic suggest positive imagery (eg, “smiling” and “nature”) while the tags of the *Suicide & self-harm* topic suggest negative imagery (eg, “dark” and “cloudy”). The *Cancer* topic contains imagery associated specifically with breast cancer awareness, with the tag “pink” appearing as an image tag, as well as in the top hashtags and caption words.

A common theme across topics is that images frequently contain people. The image tag “posing” is the top tag associated with 11 topics (ie, *Chronic illness*, *Diabetes*, *Gym/fitness*, *Health care*, *Nursing*, *Hospitalization*, *Mental health*, *Skin health*, *Tanning*, *Cosmetic surgery*, *Dental health*). Other topics have the top tag of “person” (ie, *Illness*), “group” (ie, *Cancer*), “woman” (ie, *Headaches & body aches*), and “man” (ie, *Bodybuilding*). The image tags “swimsuit” and “underwear” are especially common in many of the exercise and fitness topics. This matches an observation in a previous study of fitness images on Instagram which found that “most images contained posed individuals with some degree of objectification” [38].

Gender associations can also be observed in the image features. There are four topics that contain a male-associated image tag (ie, “man,” “boy,” “male”) but no female tag (ie, *Diabetes*, *Massage*, *Gym/fitness training*, *Bodybuilding*), and 6 topics that contain a female-associated tag (ie, “woman,” “girl,” “female”) but no male tag (ie, *Cancer*, *Musculoskeletal pain*, *Headaches & body aches*, *Exercise*, *Gym/bodybuilding*, *Gym/fitness*). Only 1 topic (ie, *Allergies*) included both genders in the top 10 image tags (ie, “woman” and “boy”).

In one case, 2 topics with very similar text features had different gender patterns in the images. Consider the topic with top hashtags, “workout,” “fitness,” “gym,” “fit,” “exercise,” and top caption words, “week,” “day,” “workout,” “work,” “good;” and a similar topic with hashtags, “workout,” “abs,” “gym,” “muscle,” “chest,” and caption words, “work,” “back,” “legs,” “leg,” “strong.” These top words do not explicitly contain gendered words, but in the image tags, the former contains “woman” while the latter contains “man” and “male.” Table 3 shows the 10 topics with the highest average proportions in documents, and Table 4 shows the cumulative proportions of each of the 10 categories.

Table 3. The 10 most prevalent individual topics in the dataset, ranked by their average topic proportion out of 150 total topics. When multiple topics have the same name, we show the integer index of the topic in parentheses to distinguish them.

Topic	Average Probability
Suicide & self-harm	0.012
Bodybuilding (Topic 135)	0.010
Exercise	0.009
Healthy food (Topic 2)	0.009
Gym/bodybuilding (Topic 14)	0.009
Marijuana	0.008
Healthy food (Topic 67)	0.008
Vitamins & supplements	0.008
Skin health	0.008
Gym/fitness	0.008

Table 4. The topic categories ranked by prevalence, where each category's prevalence is defined by the sum of the individual topic proportions of the category's topics.

Topic Category	Cumulative Probability
Diet	.086
Exercise	.076
Musculoskeletal health & dermatology	.046
Alternative medicine	.042
Chronic illness & pain	.039
Health care & medicine	.033
Mental health	.026
Substance use	.021
Sleep	.013
Acute illness	.012

Topics about diet and exercise are by far the most prevalent topics, while topics about acute illness and sleep are uncommon. Topics with high probabilities tended to be more coherent than low-probability topics. Of the topics with the highest probabilities, 24/25 (96%) were labeled as coherent by the annotators. Of the topics with the lowest probabilities, only 15/25 (60%) were labeled as coherent. The variability in average probabilities was low; the values ranged from .005 to .012.

Table 5 shows the 10 pairs of health topics with the highest correlations. Some of the strongest correlations are with the *Vitamins & supplements* topic, which co-occurs with a variety of other health topics. Table 6 shows the most correlated 10 topic pairs such that one topic is a health topic and the other is a nonhealth topic. Inspirational and supportive topics (ie, *Inspiration* and *Poetry & quotes*) tend to co-occur with mental health and exercise topics, and topics about religion and spirituality tend to co-occur with certain health topics, like *Alternative medicine* and *Yoga*.

Table 5. The 10 most correlated pairs of health topics.

Topic pair		
Topic A	Topic B	Pearson Correlation
Vitamins & supplements	Energy & hydration	.138
Vitamins & supplements	Health science	.131
Vitamins & supplements	Headaches & body aches	.121
Energy & hydration	Headaches & body aches	.067
Vitamins & supplements	Skin health	.060
Chronic illness	Mental health	.057
Chronic illness	Hospitalization	.050
Alternative medicine	Health science	.050
Running & cardio	Injuries & rehabilitation	.049
Headaches & body aches	Massage	.047

Table 6. The 10 most correlated pairs of topics, where each pair contains 1 health topic and 1 nonhealth topic.

Topic pair		
Topic A	Topic B	Pearson Correlation
Suicide & self-harm	Grunge/emo	.107
Mental health	Inspiration	.081
Hospitalization	Cats	.080
Vaccination	Politics	.063
Hospitalization	Religion/Christianity	.059
Yoga	Spirituality	.056
Alternative medicine	Spirituality	.048
Fitness training	Sexuality	.048
Suicide & self-harm	Poetry & quotes	.043
Gym/fitness	Inspiration	.036

Discussion

Principal Findings

The topic model results show a large and diverse set of health topics are discussed in Instagram. Qualitatively, we find that the top hashtags tend to be the best descriptors of topics, while caption words give some indication of what kind of messages are associated with the topics, such as whether they are more informational or conversational. The extracted image tags are generally much less coherent, though they do help characterize the types of images that are associated with each topic. For example, many of the topics related to pain contain images of animals, perhaps because users post cheerful images in response to pain. The tag “posing” appears in some topics, suggesting these posts may be informational rather than personal. The *Cancer* topic contains the image tags “group” and “crowd;” it appears these many posts in this topic are about cancer awareness events. In some cases, image tags were the defining characteristics that distinguished clusters that were otherwise very similar, which suggests that images are informative beyond the hashtags and captions to conduct content analyses of Instagram posts.

Qualitatively, it appears that in most cases the image tags are not specific enough to be useful for directly identifying posts relevant to a specific health application. However, tags of food and beverages appear to be fairly specific and accurate, suggesting that computer vision may help in identifying posts for studies of diet and food consumption. The only previous work we are aware of that used automatically extracted image tags for this purpose is [24], which found that image tags were predictive of lifestyle factors; for example, “glass”, “liquid” and “beverage” were associated with alcohol consumption. The authors suggested that image tags may be useful for identifying stigmatizing behaviors, where social media users may post images of an activity but not explicitly tag the activity. Even nonstigmatized activities, like general food consumption, may not be tagged by a user in a way that is specific enough to identify by text search, while computer vision may help. We observed that image tags extracted from the computer vision API did not usually identify a specific dish, but could at least identify broad categories, like “meat” and “vegetable,” and in some cases were more specific, like “potato” and “doughnut.” We, therefore, argue that this type of computer vision tool can expand the amount of data available for studying patterns in food consumption.

We gained additional insights by considering the co-occurrences of health topics in the data. For example, the *Vitamins & supplements* topic is less likely to appear in a post in isolation but instead co-occurs with other topics, likely because supplements are discussed in the context in which they are used. Using this data to study nutrition in a population may, therefore, be able to show how nutrition is discussed and applied to specific aspects of health. Some pairs of topics with high correlations may indicate comorbidities, such as *Chronic illness* and *Mental health* [39].

Co-occurrences with nonhealth topics may give insights into other contexts in which health is discussed. We observed that many health topics frequently co-occur with inspirational topics, such as topics containing poetry and quotes, or topics about nature, as well as topics related to spirituality and religion. These types of posts may give insight into how individuals cope with and support others with, illness and disease.

An additional observation in some topics is the use of certain hashtags to identify a specific community of users [40], such as the #wlscommunity in the weight loss topics. Online health communities have been studied to understand social support and behavior change in managing health conditions [41]. Instagram-based communities may be a unique source for studying similar issues. Communities for specific demographic groups (eg, #girlswholift) are also present. In some cases, demographic associations could be gleaned from the image tags, even if the text tags were not explicitly gendered.

One of our methodological contributions was to repurpose an existing tool, the polylingual topic model, for a new task of combining different modalities of data in a topic model. We showed that automatically extracted image tags from a computer vision API can be treated as text tokens in an existing topic model. Beyond topic models, our observations of the results suggest that these extracted image tags are in some cases useful descriptors of images. We suggest that this type of tool can be applied to images for health research more broadly.

We observed the same broad set of topics in Instagram that have previously been seen in Twitter [3], suggesting that Instagram could serve as a potential data source for many of the same applications for which Twitter is used. Moreover, the presence of the first-person language (eg, “i’ve”) in some topics indicates that health posts on Instagram include personal health mentions, which is an essential characteristic for some types of surveillance [42]. This has implications for social media-based health surveillance because this suggests that Instagram could be used as a data source for similar areas of research, while having the potential benefit of covering a larger population than Twitter. We do not suggest that one platform is universally better than another, but instead, using data from multiple platforms can result in better surveillance than reliance on one platform [43,44].

Instagram may complement Twitter as a data source because it has a different demographic distribution. The user base of Instagram is younger, lower income, and more urban compared to Twitter, [45]. These demographics cover populations that are traditionally harder to reach in health research [46], and so Instagram may be well-suited for studying such populations. This argument has been made for using Twitter [47], yet Instagram has an even heavier bias toward these populations. Additionally, Instagram has a gender bias that Twitter does not have, being nearly 50% more popular among women than men [45].

Limitations

Not all health topics are discussed widely on Instagram, which may be a limitation of using Instagram. By far the most common topics in Instagram are related to diet and exercise, while topics on acute illness, which would be needed for a task like influenza surveillance, are the least common. This may explain why all prior work we identified using image-sharing platforms for health research was related to lifestyle factors, such as diet (most common) and physical activity. Nonetheless, topics about infectious disease do exist on Instagram, and so it may be worth investigating the utility of contributing this data to an ensemble surveillance system [43]. To the best of our knowledge, no prior work has studied Instagram for infectious disease surveillance, which would be a good candidate for future research. However, this study did not collect Instagram posts from a large enough span of time to validate the data for such a task.

Another limitation of using Instagram is the limited availability of metadata. When crawling Instagram, it is difficult to sample data uniformly across time, as Instagram does not provide a streaming API analogous to Twitter’s widely-used APIs, which would make it difficult to extract the long-term pattern, for example, to validate influenza tracking. Location data also appears to be difficult to obtain when crawling from the web. In our dataset, 46% of posts contained a user-specified location string, but these were not in a standard format, and many of them were names of businesses or other specific locations, without reference to a geographic area. Geolocation from Instagram is less well understood in social media research, as well as inference of other demographic attributes that may be important in public health research. Richer data may be available from certain resellers; for example, Gnip, who is the official seller of Twitter data, also sells Instagram data, which can be searched by either tag or geolocation.

In addition to limitations of this data, there are limitations with the topic model methodology. Topic model evaluation is notoriously difficult [35], though research has found that this methodology can provide overlapping insights with more traditional, manual text analysis [48]. There is subjectiveness in choosing the number of topics and labeling and categorizing the topics, which we mitigated by having two researchers involved in each step. An advantage of the topic modeling approach for this study is that it can be applied to the entire dataset of nearly 100,000 posts, and the word distributions highlight the features associated with each topic across the three modalities. Furthermore, with such a large number of topics in the data (ie, 47 health topics identified by the approach used here), a typical sample size for manual content analysis, on the order of 1,000 posts [19-21], would be insufficient for accurately learning the prevalence of each topic.

Another limitation of topic modeling is that the topics characterize *what* is being discussed, but it is difficult to describe *how* the content is presented. For example, the topic model can identify posts that are related to marijuana, but it does not distinguish between personal marijuana use, information about marijuana, or advertisements for cannabis products—distinctions that have been made in prior work using more qualitative methods [20]. However, the topic model is still an essential first step of filtering and retrieval, after which topic-specific posts could be analyzed in more depth.

We note that there exist other methods for identifying thematic patterns in text beyond probabilistic topic models that have been used in health research, such as network-based clustering on term co-occurrence graphs [49]. Most such methods, including topic models, rely on co-occurrence statistics of words and have similar properties and limitations. We used the polylingual topic model due to its ability to integrate different “languages” or modalities of data.

Finally, the grouping of topics into ten overlapping categories is also limited. Some topics were difficult to categorize, and the boundaries of some categories were difficult to define. However, the goal of the categorization is to present the raw results (available from our dataset in Multimedia Appendix 1) more concisely. The mapping of topics to categories is transparent (viewable in Multimedia Appendix 2) so that the results can be interpreted correctly.

Conclusion

This study shows that health is discussed on Instagram in a variety of ways, and there is potential for computer vision techniques to automatically characterize health-related images, which could extend public health surveillance of social media beyond text-based analysis. Our dataset of nearly 100,000 posts is available to allow for the study of specific topics and image tags in more depth. There are pragmatic reasons why this popular platform has been used in research relatively little compared to platforms like Twitter and Facebook, but our results and discussion point to ideas that image-sharing platforms like Instagram may complement other social media data sources in health research.

Conflicts of Interest

MJP serves on the advisory board to Sickweather, a company that uses social media to forecast illness.

Multimedia Appendix 1

The dataset of 96,426 Instagram posts. The raw data is not included for privacy reasons, but can be collected through the URLs provided. The dataset includes the additional information inferred for each post: the image tags, and the topic model probabilities. Descriptions of all 150 topics are also included, as well as the 269 keywords used to search for posts.

[ZIP File (Zip Archive), 5MB - publichealth_v4i2e10150_app1.zip]

Multimedia Appendix 2

The descriptions of the 47 health topics. Each slide contains a table corresponding to one of the ten health categories, along with the top ten hashtag, caption, and image features for each topic in that category. Some topics have the same name, in which case we added the topic index (1-150) in parentheses to the name to differentiate these topics.

[PDF File (Adobe PDF File), 267KB - publichealth_v4i2e10150_app2.pdf]

References

1. Paul MJ, Dredze M. Social Monitoring for Public Health. In: Synthesis Lectures on Information Concepts, Retrieval, and Services. San Rafael, California: Morgan & Claypool; 2017:1-185.
2. Statista. Number of monthly active Twitter users worldwide from 1st quarter to 3rd quarter. 2010. 2018 URL: https://www.statista.com/statistics/282087/number-of-monthly-active-twitter-users/ [accessed 2018-02-08] [WebCite Cache ID 6x53kEUB1]
3. Paul MJ, Dredze M. Discovering health topics in social media using topic models. PLoS One 2014;9(8):e103408 [FREE Full text] [doi: 10.1371/journal.pone.0103408] [Medline: 25084530]
4. Signorini A, Segre AM, Polgreen PM. The use of Twitter to track levels of disease activity and public concern in the U.S. during the influenza A H1N1 pandemic. PLoS One 2011;6(5):e19467 [FREE Full text] [doi: 10.1371/journal.pone.0019467] [Medline: 21573238]
5. Sadilek A, Kautz H, Silenzio V. Modeling spread of disease from social interactions. 2012 Presented at: Sixth AAAI International Conference on Weblogs and Social Media; 2012; Dublin, Ireland.
6. Fried D, Surdeanu M, Kobourov S, Hingle M, Bell D. Analyzing the language of food on social media. 2014 Presented at: IEEE International Conference on Big Data; 2014; Washington, DC.

7. Nguyen QC, Li D, Meng H, Kath S, Nsoesie E, Li F, et al. Building a National Neighborhood Dataset From Geotagged Twitter Data for Indicators of Happiness, Diet, and Physical Activity. JMIR Public Health Surveill 2016 Oct 17;2(2):e158 [FREE Full text] [doi: 10.2196/publichealth.5869] [Medline: 27751984]
8. Hanson CL, Burton SH, Giraud-Carrier C, West JH, Barnes MD, Hansen B. Tweaking and tweeting: exploring Twitter for nonmedical use of a psychostimulant drug (Adderall) among college students. J Med Internet Res 2013;15(4):e62 [FREE Full text] [doi: 10.2196/jmir.2503] [Medline: 23594933]
9. Seaman I, Giraud-Carrier C. Prevalence and attitudes about illicit and pre- scription drugs on Twitter. 2016 Presented at: IEEE International Conference on Healthcare Informatics; 2016; Chicago, IL.
10. Ordun C, Blake JW, Rosidi N, Grigoryan V, Reffett C, Aslam S, et al. Open source health intelligence for foodborne illness event characterization. Online Journal of Public Health Informatics 2013;5(1):e128.
11. Sadilek A, Kautz H, DiPrete L, Labus B, Portman E, Teitel J, et al. Deploying nEmesis: Preventing foodborne illness by data mining social media. 2016 Presented at: Twenty-Eighth Annual Conference on Innovative Applications of Artificial Intelligence; 2016; Phoenix, AZ.
12. De CM, Gamon M. Predicting depression via social media. 2013 Presented at: International Conference on Weblogs and Social Media; 2013; Boston, MA.
13. Coppersmith G, Dredze M, Harman C, Hollingshead K, Mitchell M. CLPsych 2015 Shared Task: Depression and PTSD on Twitter. 2015 Presented at: NAACL Workshop on Computational Linguistics and Clinical Psychology; 2015; Denver, CO.
14. Statista. Number of monthly active Instagram users from January to September. 2013. 2018 URL: https://www.statista.com/statistics/253577/number-of-monthly-active-instagram-users/ [accessed 2018-02-08] [WebCite Cache ID 6x57Ymx2Y]
15. Mejova Y, Haddadi H, Noulas A, Weber I. #foodporn: Obesity patterns in culinary interactions. 2015 Presented at: International Conference on Digital Health; 2015; Florence, Italy.
16. De CM, Sharma S. Kiciman E Characterizing dietary choices, nutrition, and language in food deserts via social media. 2016 Presented at: Conference on Computer Supported Cooperative Work and Social Computing; 2016; San Francisco, CA.
17. Yom-Tov E, Fernandez-Luque L, Weber I, Crain SP. Pro-anorexia and pro-recovery photo sharing: a tale of two warring tribes. J Med Internet Res 2012;14(6):e151 [FREE Full text] [doi: 10.2196/jmir.2239] [Medline: 23134671]
18. Pless R, Begtrup R, Alkulaib L, Counts S, Harnett J, Manning J, et al. Recognizing images of eating disorders in social media. 2017 Presented at: AMIA Workshop on Social Media Mining for Health Applications; 2017; Washington, DC.
19. Chu K, Allem J, Cruz TB, Unger JB. Vaping on Instagram: cloud chasing, hand checks and product placement. Tob Control 2016 Sep 22. [doi: 10.1136/tobaccocontrol-2016-053052] [Medline: 27660111]
20. Cavazos-Rehg PA, Krauss MJ, Sowles SJ, Bierut LJ. Marijuana-Related Posts on Instagram. Prev Sci 2016 Aug;17(6):710-720. [doi: 10.1007/s11121-016-0669-9] [Medline: 27262456]
21. Cherian R, Westbrook M, Ramo D, Sarkar U. Representations of Codeine Misuse on Instagram: Content Analysis. JMIR Public Health Surveill 2018 Mar 20;4(1):e22. [doi: 10.2196/publichealth.8144] [Medline: 29559422]
22. Allem J, Escobedo P, Chu K, Boley CT, Unger JB. Images of Little Cigars and Cigarillos on Instagram Identified by the Hashtag #swisher: Thematic Analysis. J Med Internet Res 2017 Jul 14;19(7):e255 [FREE Full text] [doi: 10.2196/jmir.7634] [Medline: 28710057]
23. Allem J, Chu K, Cruz TB, Unger JB. Waterpipe Promotion and Use on Instagram: #Hookah. Nicotine Tob Res 2017 Jan 11. [doi: 10.1093/ntr/ntw329] [Medline: 28077449]
24. Garimella VRK, Alfayad A, Weber I. Social media image analysis for public health. 2016 Presented at: CHI Conference on Human Factors in Computing Systems; 2016; San Jose, CA.
25. Wang S, Paul M, Dredze M. Exploring health topics in Chinese social media: an analysis of Sina Weibo. 2014 Presented at: AAAI Workshop on the World Wide Web and Public Health Intelligence; 2014; Québec City, Canada.
26. Instagram Developer Documentation. 2018 URL: https://www.instagram.com/developer/ [accessed 2018-05-04] [WebCite Cache ID 6zApRoL0h]
27. Twitter: Sample realtime tweets. 2018 URL: https://developer.twitter.com/en/docs/tweets/sample-realtime/overview/GET_statuse_sample [accessed 2018-05-04] [WebCite Cache ID 6zApUY2BJ]
28. Lui M, Baldwin T. Langid.py: an off-the-shelf language identification tool. 2012 Presented at: Association for Computational Linguistics; 2012; Jeju, South Korea.
29. Bird S, Loper E, Klein E. Natural Language Processing with Python. Sebastopol, CA: O'Reilly Media Inc; 2009.
30. Microsoft A. Computer Vision API. 2018 URL: https://azure.microsoft.com/en-us/services/cognitive-services/computer-vision/ [accessed 2018-02-08] [WebCite Cache ID 6x57iYvRl]
31. Blei DM. Probabilistic topic models. Commun. ACM 2012 Apr 01;55(4):77. [doi: 10.1145/2133806.2133826]
32. Prier K, Smith M, Giraud-Carrier C, Hanson C. Identifying health-related topics on Twitter: An exploration of tobacco-related tweets as a test topic. 2011 Presented at: International Conference on Social Computing, Behavioral-cultural Modeling and Prediction; 2011; College Park, MD.

33. Ghosh DD, Guha R. What are we 'tweeting' about obesity? Mapping tweets with Topic Modeling and Geographic Information System. Cartogr Geogr Inf Sci 2013;40(2):90-102 [FREE Full text] [doi: 10.1080/15230406.2013.776210] [Medline: 25126022]
34. Wallace BC, Paul MJ, Sarkar U, Trikalinos TA, Dredze M. A large-scale quantitative analysis of latent factors and sentiment in online doctor reviews. J Am Med Inform Assoc 2014;21(6):1098-1103 [FREE Full text] [doi: 10.1136/amiajnl-2014-002711] [Medline: 24918109]
35. Chang J, Boyd-Graber J, Gerrish S, Wang C, Blei D. Reading tea leaves: how humans interpret topic models. 2009 Presented at: International Conference on Neural Information Processing Systems (NIPS); 2009; Vancouver, Canada.
36. Mimno D, Wallach H, Naradowsky J, Smith DA, McCallum A. Polylingual topic models. 2009 Presented at: Empirical Methods in Natural Language Processing; 2009; Singapore.
37. McCallum AK. 2002. MALLET: A Machine Learning for Language Toolkit URL: http://mallet.cs.umass.edu/ [accessed 2018-06-18] [WebCite Cache ID 70HBNf8dX]
38. Santarossa S, Coyne P, Lisinski C, Woodruff SJ. #fitspo on Instagram: A mixed-methods approach using Netlytic and photo analysis, uncovering the online discussion and author/image characteristics. J Health Psychol 2016 Nov 01:1359105316676334. [doi: 10.1177/1359105316676334] [Medline: 27852889]
39. Fishbain DA, Cutler R, Rosomoff HL, Rosomoff RS. Chronic pain-associated depression: antecedent or consequence of chronic pain? A review. Clin J Pain 1997 Jun;13(2):116-137. [Medline: 9186019]
40. Cook J, Kenthapadi K, Mishra N. Group chats on Twitter. 2013 Presented at: International Conference on World Wide Web; 2013; Rio de Janeiro, Brazil.
41. Willis E, Royne MB. Online Health Communities and Chronic Disease Self-Management. Health Commun 2017 Mar;32(3):269-278. [doi: 10.1080/10410236.2016.1138278] [Medline: 27218836]
42. Karisani P, Agichtein E. Did you really just have a heart attack? Towards robust detection of personal health mentions in social media. 2018 Presented at: The Web Conference (WWW); 2018; Lyon, France.
43. Santillana M, Nguyen AT, Dredze M, Paul MJ, Nsoesie EO, Brownstein JS. Combining Search, Social Media, and Traditional Data Sources to Improve Influenza Surveillance. PLoS Comput Biol 2015 Oct;11(10):e1004513 [FREE Full text] [doi: 10.1371/journal.pcbi.1004513] [Medline: 26513245]
44. McGough SF, Brownstein JS, Hawkins JB, Santillana M. Forecasting Zika Incidence in the 2016 Latin America Outbreak Combining Traditional Disease Surveillance with Search, Social Media, and News Report Data. PLoS Negl Trop Dis 2017 Jan;11(1):e0005295 [FREE Full text] [doi: 10.1371/journal.pntd.0005295] [Medline: 28085877]
45. Greenwood S, Perrin A, Duggan M. 2016. Pew Research Center: Social media update 2016 URL: http://www.pewinternet.org/2016/11/11/social-media-update-2016/ [WebCite Cache ID 6yMIOCpgJ]
46. Bonevski B, Randell M, Paul C, Chapman K, Twyman L, Bryant J, et al. Reaching the hard-to-reach: a systematic review of strategies for improving health and medical research with socially disadvantaged groups. BMC Med Res Methodol 2014;14:42 [FREE Full text] [doi: 10.1186/1471-2288-14-42] [Medline: 24669751]
47. Huang X, Smith M, Paul M, Ryzhkov D, Quinn S, Broniatowski D, et al. Examining patterns of influenza vaccination in social media. 2017 Presented at: AAAI Joint Workshop on Health Intelligence; 2017; San Francisco, CA.
48. Baumer E, Mimno D, Guha S, Quan E, Gay G. Comparing grounded theory and topic modeling: Extreme divergence or unlikely convergence? Journal of the Association for Information Science and Technology 2017 Apr 28;68(6):1397-1410 [FREE Full text] [doi: 10.1002/asi.23786]
49. Allem J, Ferrara E, Uppu SP, Cruz TB, Unger JB. E-Cigarette Surveillance With Social Media Data: Social Bots, Emerging Topics, and Trends. JMIR Public Health Surveill 2017 Dec 20;3(4):e98 [FREE Full text] [doi: 10.2196/publichealth.8641] [Medline: 29263018]

Abbreviations

API: Application Programming Interface
NLTK: Natural Language Toolkit

Edited by T Sanchez; submitted 15.02.18; peer-reviewed by JP Allem, N O'Kane, Y Mejova; comments to author 29.03.18; revised version received 17.05.18; accepted 29.05.18; published 29.06.18.

Please cite as:
Muralidhara S, Paul MJ
#Healthy Selfies: Exploration of Health Topics on Instagram
JMIR Public Health Surveill 2018;4(2):e10150
URL: http://publichealth.jmir.org/2018/2/e10150/
doi:10.2196/10150
PMID:29959106

Original Paper

HIV Clustering in Mississippi: Spatial Epidemiological Study to Inform Implementation Science in the Deep South

Thomas J Stopka[1,2], MHS, PhD; Lauren Brinkley-Rubinstein[3,4], PhD; Kendra Johnson[5], MPH; Philip A Chan[6,7], MS, MD; Marga Hutcheson[1], MA; Richard Crosby[8], PhD; Deirdre Burke[1], MPH; Leandro Mena[9], MD, MPH; Amy Nunn[6,7,10], MS, ScD

[1]Department of Public Health and Community Medicine, School of Medicine, Tufts University, Boston, MA, United States
[2]Clinical and Translational Science Institute, Tufts University School of Medicine, Boston, MA, United States
[3]Department of Social Medicine, University of North Carolina, Chapel Hill, NC, United States
[4]Center for Health Equity Research, University of North Carolina, Chapel Hill, NC, United States
[5]Mississippi State Department of Health, Jackson, MS, United States
[6]School of Public Health, Brown University, Providence, RI, United States
[7]Department of Medicine, Brown University, Providence, RI, United States
[8]College of Public Health, University of Kentucky, Lexington, KY, United States
[9]John D Bower School of Population Health, University of Mississippi Medical Center, Jackson, MS, United States
[10]Rhode Island Public Health Institute, Providence, RI, United States

Corresponding Author:
Thomas J Stopka, MHS, PhD
Department of Public Health and Community Medicine
School of Medicine
Tufts University
136 Harrison Ave
MV244
Boston, MA, 02111
United States
Phone: 1 617 636 2110
Fax: 1 617 636 4017
Email: thomas.stopka@tufts.edu

Abstract

Background: In recent years, more than half of new HIV infections in the United States occur among African Americans in the Southeastern United States. Spatial epidemiological analyses can inform public health responses in the Deep South by identifying HIV hotspots and community-level factors associated with clustering.

Objective: The goal of this study was to identify and characterize HIV clusters in Mississippi through analysis of state-level HIV surveillance data.

Methods: We used a combination of spatial epidemiology and statistical modeling to identify and characterize HIV hotspots in Mississippi census tracts (n=658) from 2008 to 2014. We conducted spatial analyses of all HIV infections, infections among men who have sex with men (MSM), and infections among African Americans. Multivariable logistic regression analyses identified community-level sociodemographic factors associated with HIV hotspots considering all cases.

Results: There were HIV hotspots for the entire population, MSM, and African American MSM identified in the Mississippi Delta region, Southern Mississippi, and in greater Jackson, including surrounding rural counties ($P<.05$). In multivariable models for all HIV cases, HIV hotspots were significantly more likely to include urban census tracts (adjusted odds ratio [AOR] 2.01, 95% CI 1.20-3.37) and census tracts that had a higher proportion of African Americans (AOR 3.85, 95% CI 2.23-6.65). The HIV hotspots were less likely to include census tracts with residents who had less than a high school education (AOR 0.95, 95% CI 0.92-0.98), census tracts with residents belonging to two or more racial/ethnic groups (AOR 0.46, 95% CI 0.30-0.70), and census tracts that had a higher percentage of the population living below the poverty level (AOR 0.51, 95% CI 0.28-0.92).

Conclusions: We used spatial epidemiology and statistical modeling to identify and characterize HIV hotspots for the general population, MSM, and African Americans. HIV clusters concentrated in Jackson and the Mississippi Delta. African American race and urban location were positively associated with clusters, whereas having less than a high school education and having a higher percentage of the population living below the poverty level were negatively associated with clusters. Spatial epidemiological analyses can inform implementation science and public health response strategies, including improved HIV testing, targeted prevention and risk reduction education, and tailored preexposure prophylaxis to address HIV disparities in the South.

(JMIR Public Health Surveill 2018;4(2):e35) doi:10.2196/publichealth.8773

KEYWORDS

hotspots; HIV; racial disparities; social determinants of health; HIV treatment; HIV screening

Introduction

In the United States, HIV infections cluster geographically [1,2]. New HIV infections are concentrated in the Southeast and in neighborhoods with high rates of poverty that are predominately Hispanic/Latino or African American [2,3]. Geographic information system (GIS) and spatial epidemiological analyses provide opportunities to better understand the spatial distribution of HIV/AIDS, high-risk areas for disease transmission and acquisition [4], access to prevention interventions [5], and “hotspot” clustering of HIV-related mortality [6]. Understanding geographic clustering of HIV infections can also be helpful for directing limited public health resources to communities that bear a disproportionate share of HIV disease burden. Spatial epidemiological and geostatistical analyses can facilitate identification and characterization of HIV hotspot clusters down to the neighborhood level [6], which can be useful for targeting HIV prevention and care interventions in communities with high rates of HIV infection, and specific sociodemographic characteristics associated with clustering.

The southern United States accounted for 52% of HIV infections in 2014 [7]. Mississippi ranks ninth highest in rate of new HIV infections [2] and had the eighth-highest AIDS death rate of any state in the United States in 2016 [8]. Mississippi has alarming sociodemographic and geographical disparities related to HIV/AIDS; African Americans comprise only approximately 38% of the total population, but accounted for 80% of diagnosed HIV cases in 2014 [9]. Additionally, Jackson, Mississippi, had the fifth-highest AIDS diagnosis rate in 2014 and ranked fourth for HIV infection among metropolitan areas in the United States. Moreover, from 2005 to 2014, the number of infections among men who have sex with men (MSM) in Mississippi increased 59%, with an even sharper increase among young African American MSM [9,10]. A recent study found that Jackson has the highest rate of new HIV infections among MSM of any city nationwide [11]. Further, a recently completed study of 609 young African American MSM residing in Jackson observed that 27.9% were HIV infected at baseline and 6.8% of the remainder acquired HIV during the ensuing 12 months [12].

The goal of this study was to identify and characterize HIV hotspot clusters in Mississippi. We aimed to test the hypotheses that HIV infections clustered geospatially in Mississippi and that HIV hotspots were associated with sociodemographic factors (eg, race, income, education) in local communities.

Methods

Data Sources

We obtained deidentified HIV surveillance data from the Mississippi State Department of Health. The dataset included newly reported HIV cases from 2008 to 2014 (N=3410) with variables such as race, ethnicity, age at diagnosis, year of diagnosis, sex at birth, sexual orientation, and the census tract of residence at the time of diagnosis. The HIV cases were geocoded by the Mississippi State Department of Health. Among the 3410 reported cases, 2732 (80.12%) had a census tract of residence identified (ie, geocoded). In comparative analyses, we did not find any significant differences between reported cases that had a census tract of residence versus those that did not. Cases identified in prisons and decedent cases were excluded from the dataset, resulting in a total of 2048 cases. These 2048 georeferenced cases were aggregated at the census tract level to facilitate spatial analyses and creation of GIS maps, while protecting the confidentiality of people living with HIV.

We also obtained population denominators and community-level sociodemographic data, on the census tract level, from the US Census Bureau’s American Community Survey (ACS) 2010-2014 Five-Year Estimates [13]. In the 2010 Census, there were 664 census tracts in Mississippi. Using the surveillance and ACS data, we calculated HIV rates per 100,000 people at the census tract level. Census tracts with a total population of less than one were deemed as “uninhabited” and were excluded from all analyses, resulting in an omission of six tracts in Mississippi. Our final analytical dataset included HIV rates and community-level covariates for 658 Mississippi census tracts.

Geographic Information System Mapping and Spatial Analysis

First, we created descriptive GIS maps to determine the initial spatial distribution of HIV cases and rates across all Mississippi census tracts. We categorized the HIV counts and rates by quintiles in thematic maps to create cut points at the 20th, 40th, 60th, and 80th percentiles. The lowest quintile represented HIV levels from zero to the 20th percentile (lowest HIV burden) and the highest quintile represented HIV levels for census tracts that fell between the 80th to 100th percentile (highest HIV burden). Initial descriptive maps for distinct study years did not depict substantial variation across years. Aggregated data across all study years also provided greater statistical power to identify clusters. Because we had access to data for our outcomes and covariates at the census tract level (rather than the address level), we opted to use a five-step spatial analytical approach to identify HIV hotspots:

1. First, we temporarily excluded large census tracts (outliers that were >1.5 standard deviations above the mean square mile area for all census tracts in Mississippi) that might bias distance calculations in subsequent steps.
2. Next, we calculated the spatial connectivity of the census tracts in Mississippi by calculating the mean and maximum distance between the geocentroid (ie, the geographic center) of each census tract and the geocentroids of the two nearest neighboring census tracts to obtain distance parameters for the next step.
3. We conducted incremental spatial autocorrelation analyses to determine the distance at which clustering for our outcome of interest was most intense (ie, we identified the most significant spatial sphere of influence with regard to HIV clustering in Mississippi) [14].
4. We calculated a spatial weights matrix that accounted for the spatial relationships of all census tracts, including the large census tracts excluded from previous steps, and our HIV measures (ie, counts and rates), improving the validity of the hotspot analysis.
5. We calculated the Getis-Ord Gi* statistic, which produces *z* scores, to identify clustering patterns across all Mississippi census tracts. Ultimately, a census tract was identified as belonging to a hotspot (or coldspot) cluster when it, and its neighboring census tracts, had a local mean HIV count/rate that was higher or lower than the mean HIV rate for all census tracts in Mississippi (ie, when the local mean HIV rate was higher or lower than the global mean HIV rate).

We have described hotspot analyses in more detail elsewhere [15,16]. All maps and spatial analyses were conducted in ArcGIS v10.3.1 (ESRI, Redlands, CA, USA).

Statistical Analyses

We assessed measures of central tendency (ie, mean, median, 95% confidence interval, and interquartile ranges) for all community-level sociodemographic explanatory variables. We calculated these descriptive statistics for all census tracts in Mississippi (n=658), census tracts located within HIV hotspot clusters (n=160), and census tracts outside of HIV hotspots (ie, tracts with mean HIV rates and coldspot clusters; n=498). Based on ACS data from the US Census Bureau, we created dichotomous categorical variables for the percentage of the population that was male and female, median annual individual income, median annual household income, percentage white, percentage African American, percentage Hispanic/Latino ethnicity, percentage Asian, percentage Hawaiian/Pacific Islander, percentage other races/ethnicities, percentage with two or more ethnicities, percentage of households on food stamps, and population density. Based on the statistical distribution of the data, we created trichotomous categorical variables with cut points at the 33rd and 66th percentiles for the percentage of population living 100% below the poverty level, families with two or more workers, and families with one person working in the household.

We used bivariate logistic regressions to assess crude associations between sociodemographic factors and HIV hotspot clusters (yes/no) on the census tract level. Associations with a $P<.25$ were included in the multivariable logistic regression models. A variance inflation factor greater than 6 indicated collinearity and led to exclusion of variables from the multivariable logistic regressions. We tested interactions for each of the race variables and the education variables: high school graduate or higher, and less than a high school education. In our final adjusted model, we considered associations with a $P<.05$ as statistically significant. Statistical analyses were conducted in SAS 9.3 (Cary, NC, USA).

Results

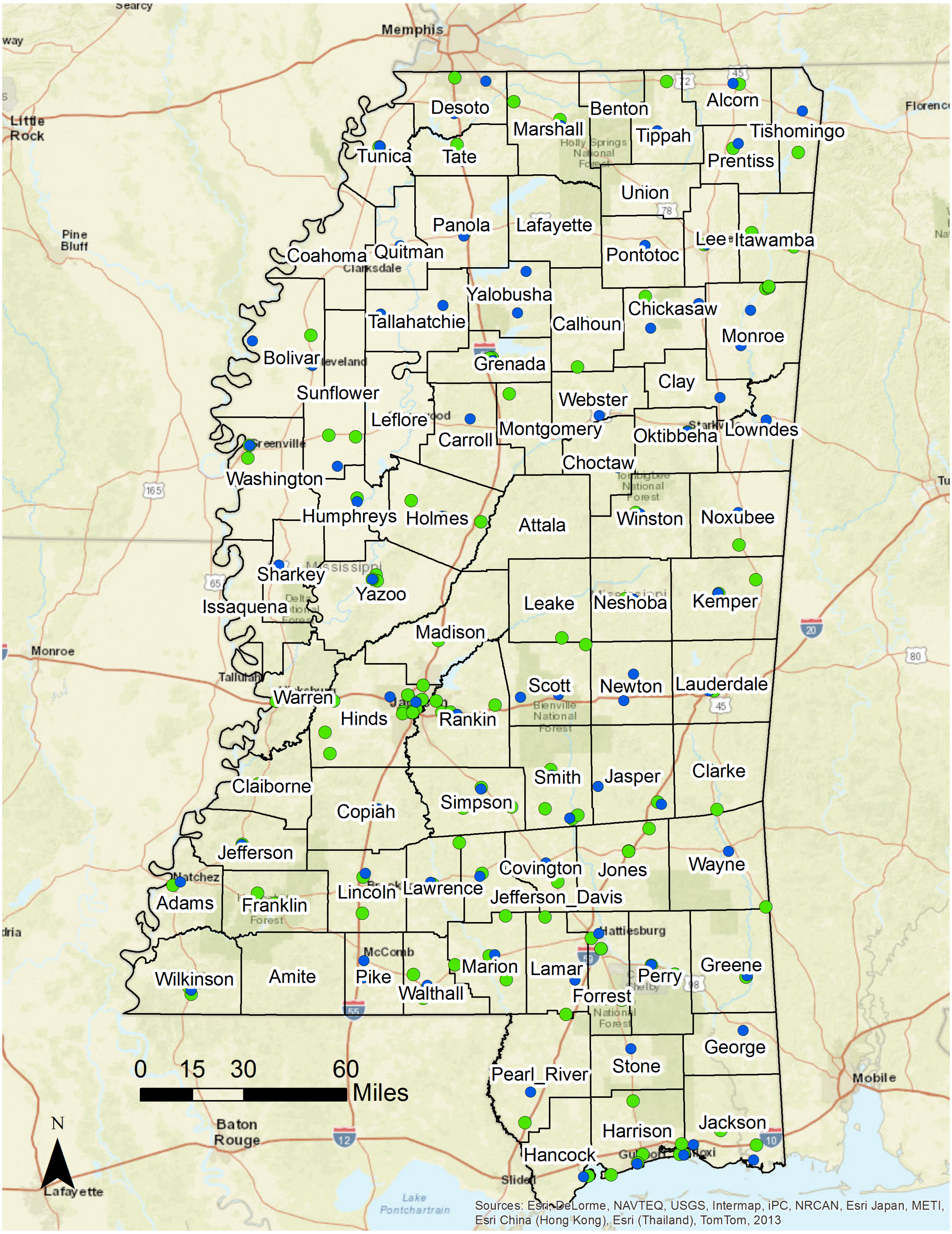
Between 2008 and 2014, 3410 HIV cases were reported to the Mississippi Department of Health HIV surveillance system. Of these, 80.12% (2732/3410) had a census tract of residence and were successfully geocoded. In comparative analyses, we did not find any significant differences between reported HIV cases that had a census tract of residence versus those that did not. Typical reasons for lack of a geocoded case were missing address, incomplete address, and inclusion of a PO box as an address. Cases identified in prisons and decedent cases were excluded from the dataset, resulting in a total of 2048 cases in our final analytical dataset. These 2048 georeferenced cases were aggregated at the census tract level (n=658) to facilitate spatial analyses and creation of GIS maps. Figure 1 includes a reference map of Mississippi that portrays county boundaries, major highways, and locations with free HIV testing services, including Ryan White clinics and County Health Departments that receive support for HIV screening from the Mississippi State Department of Health.

Our initial descriptive maps showed high HIV rates per 100,000 people dispersed throughout many regions of Mississippi, with rates of up to 127 to 1350 cases per 100,000 in the Jackson area for the entire population (Figure 2) and up to 271 to 4054 HIV cases per 100,000 among African Americans alone (Figure 3). Some census tracts in nonurban areas also had rates of HIV infection that fell between the 80th and 100th percentile (ie, the upper quintile).

Results from our hotspot cluster analyses indicated that there was a large statistically significant hotspot cluster for HIV rates in Jackson and surrounding counties, with smaller hotspots in the Mississippi Delta region, including Cleveland, Bolivar, Sunflower, Leflore, Coahoma, and Tallahatchie Counties (Figure 4). Census tracts in the greater Jackson area also appeared in HIV hotspot clusters when the analysis was limited to African Americans. Smaller HIV clusters in Pearl River and Stone Counties were also identified for African Americans (Figure 5).

In our cluster analysis focused on HIV cases among MSM, we again identified the greater Jackson area as a hotspot, and we identified smaller hotspots in Pearl River County and Greene/Wayne counties (Figure 6). When we limited our cluster analysis to only those HIV cases identified as both African American and MSM, we again identified a hotspot in the greater Jackson area (Figure 7).

Figure 1. Locations of free HIV testing services that receive support for HIV screening from the Mississippi State Department of Health, 2014. Green points represent Ryan White clinics; Blue points represent County Health Departments.



Sources: Esri, DeLorme, NAVTEQ, USGS, Intermap, iPC, NRCAN, Esri Japan, METI, Esri China (Hong Kong), Esri (Thailand), TomTom, 2013

Figure 2. HIV rates per 100,000 population in Mississippi, 2008-2014.

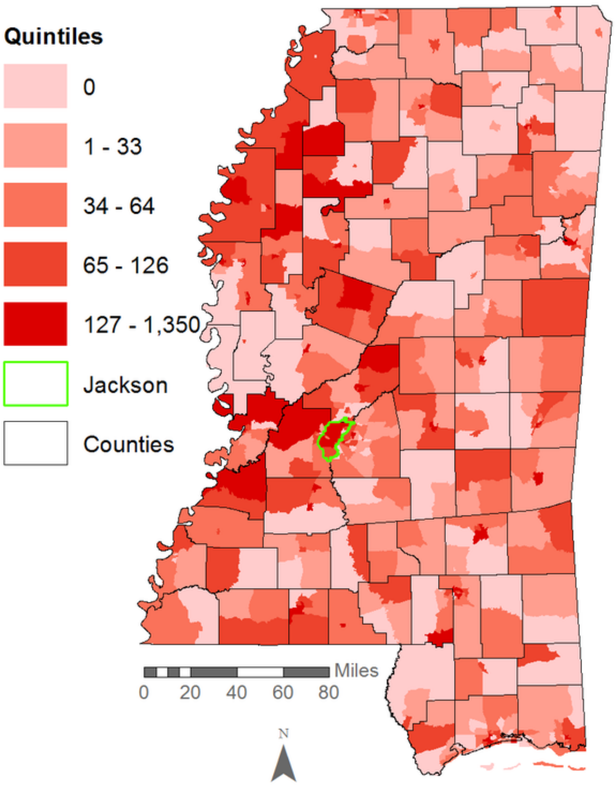


Figure 3. HIV rates per 100,000 African Americans in Mississippi, 2008-2014.

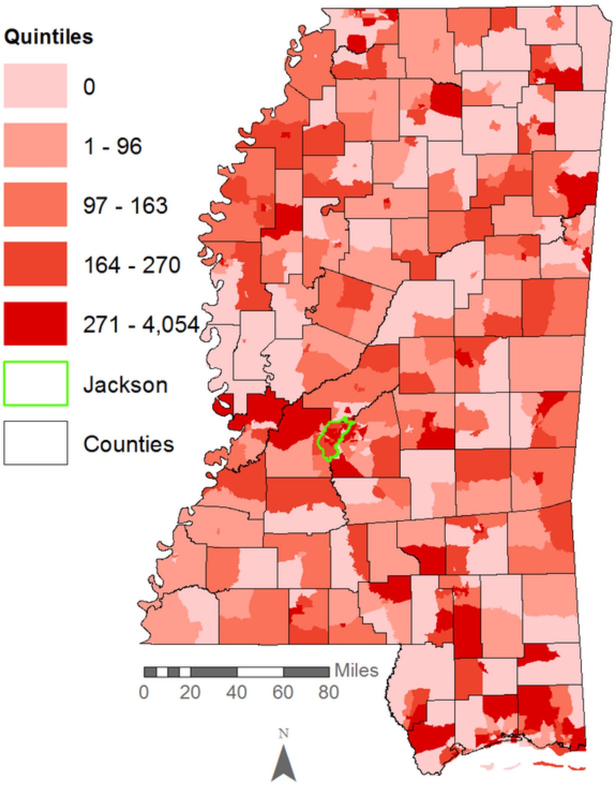


Figure 4. Hotspot cluster map for HIV rates per 100,000 population in Mississippi, 2008-2014. Clusters are based on HIV rates aggregated at the census tract level. Census tracts with elevated HIV rates (red) represent hotspots (P<.05); census tracts with low HIV rates (blue) represent coldspots (P<.05); census tracts with mean HIV rates are represented in yellow. The Jackson Metropolitan Area is outlined in green.

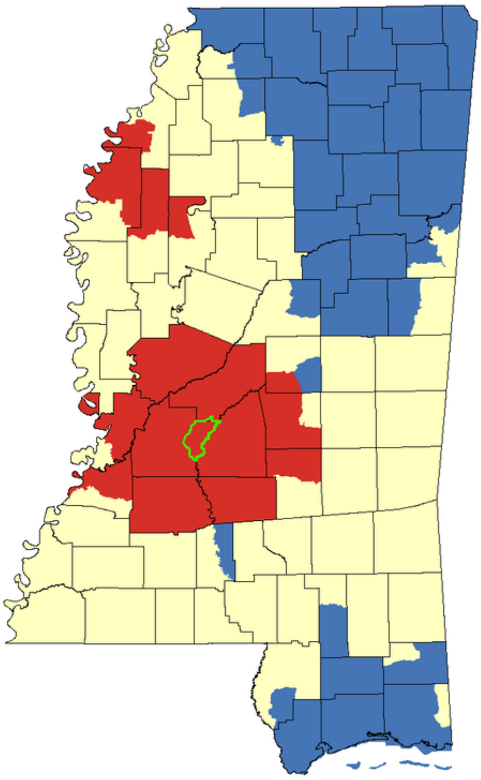


Figure 5. Hotspot cluster map for HIV rates per 100,000 African Americans in Mississippi, 2008-2014. Clusters are based on African American HIV rates aggregated at the census tract level. Census tracts with elevated African American HIV rates (red) represent hotspots (P<.05); census tracts with low African American HIV rates (blue) represent coldspots (P<.05); census tracts with mean African American HIV rates are represented in yellow. The Jackson Metropolitan Area is outlined in green.

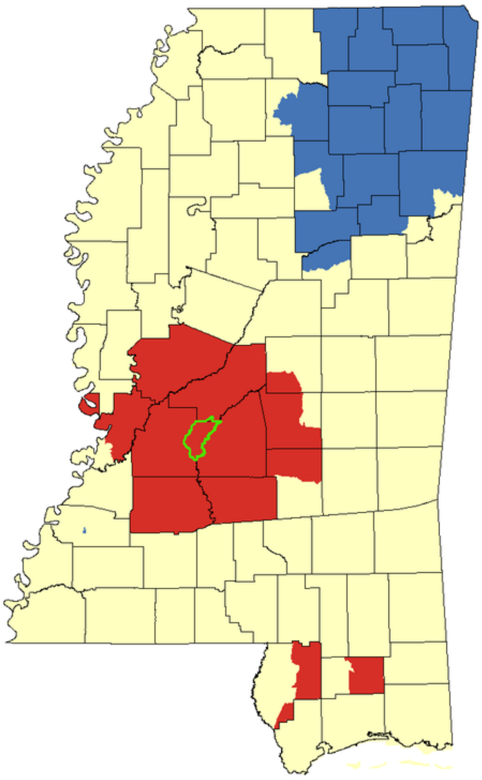


Figure 6. Hotspot cluster map for HIV case counts among MSM in Mississippi, 2008-2014. Clusters are based on MSM case counts aggregated at the census tract level. Census tracts with elevated numbers of MSM living with HIV (red) represent hotspots (P<.05); census tracts with low numbers of MSM living with HIV (blue) represent coldspots (P<.05); census tracts with mean numbers of MSM living with HIV are represented in yellow. The Jackson Metropolitan Area is outlined in green.

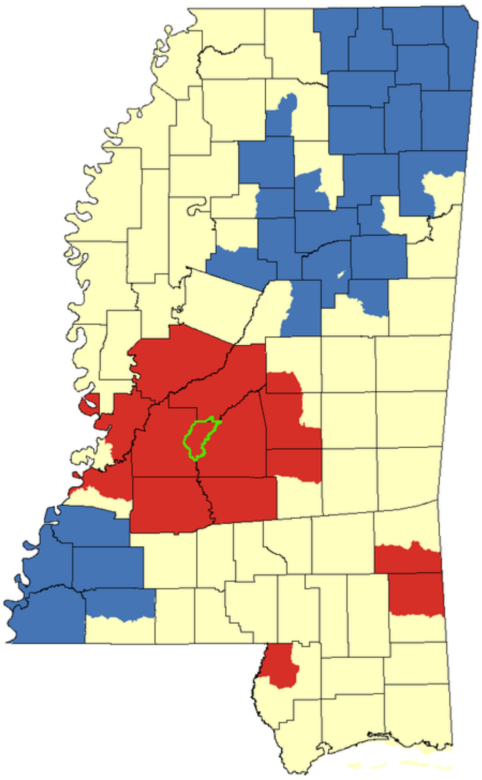


Figure 7. Hotspot cluster map for HIV case counts among African American MSM in Mississippi, 2008-2014. Clusters are based on African American MSM case counts aggregated at the census tract level. Census tracts with elevated numbers of African American MSM living with HIV (red) represent hotspots (P<.05); census tracts with low numbers of African American MSM living with HIV (blue) represent coldspots (P<.05); census tracts with mean numbers of African American MSM living with HIV counts are represented in yellow. The Jackson Metropolitan Area is outlined in green.

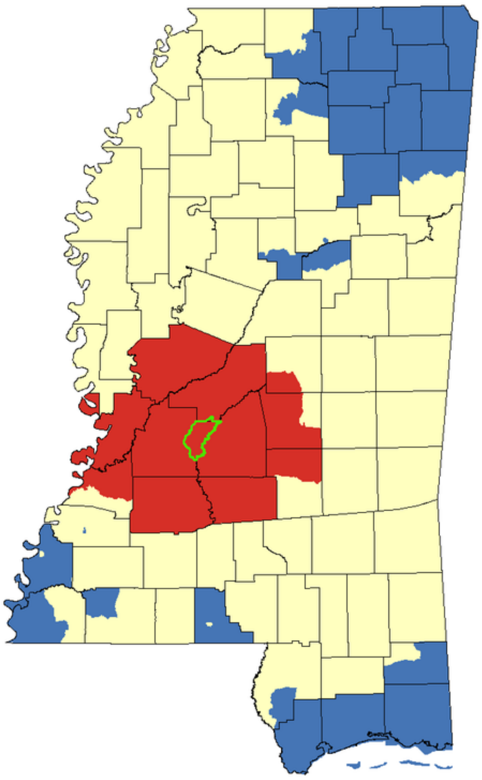


Table 1. Descriptive statistics of Mississippi census tracts, 2010-2014 (n=658).

Characteristic	Census tracts in HIV hotspots, mean (95% CI) (n=160)	Census tracts outside HIV hotspots, mean (95% CI) (n=498)
Sex (%)		
Male	48.0 (47.1-48.8)	48.78 (48.3-49.2)
Female	52.1 (51.2-52.9)	51.2 (50.8-51.7)
Total population	4410.7 (4052.4-4768.9)	4584.2 (4415.3-4753)
Median age (years)	36.1 (35.1-37.0)	37.31 (36.8-37.8)
<5	6.9 (6.5-7.3)	6.7 (6.5-6.9)
5-17	18.4 (17.6-19.1)	17.7 (17.4-18.1)
18-24	10.3 (9.4-11.3)	10.6 (10.0-11.3)
25-44	26.3 (25.4-27.4)	24.9 (24.5-25.3)
45-54	13.5 (13-14)	13.5 (13.2-13.7)
55-64	12.1 (11.5-12.7)	12.3 (12.0-12.6)
65-74	7.1 (6.6-7.5)	8.2 (7.9-8.5)
≥75	5.4 (5.1-5.8)	6.2 (6.0-6.4)
Race/Ethnicity (%)		
White	43.7 (39-48.5)	59 (56.7-61.3)
African American	53.5 (48.6-58.3)	37.6 (35.3-39.9)
Hispanic	2.4 (1.8-3.1)	3.0 (2.6-3.3)
Asian	0.98 (0.71-1.3)	0.81 (0.67-0.96)
Hawaiian/Pacific Islander	0.01 (0-0.02)	0.02 (0.01-0.03)
Other race[a]	0.94 (0.58-1.3)	0.76 (0.61-0.96)
Two or more races/ethnicities	0.66 (0.54-0.78)	1.3 (1.1-1.5)
Less than very well-spoken English (%)	1.8 (1.3-2.3)	1.5 (1.3-1.7)
Education (%)		
Less than high school education	16.9 (15.3-18.5)	20.0 (19.3-20.8)
High school graduate or higher	83.1 (75.6-84.4)	78.0 (79.2-80.7)
Median individual income (US$)	23,780 (22,274-25,286)	20,498.8 (19,970-21,027.6)
Socioeconomic measures (%)		
Below 100% of the poverty level	10.3 (8.9-11.6)	12.1 (11.3-12.9)
Population per square mile	1224 (1020-1429)	619.3 (537.1-701.4)
Own housing (%)	63.8 (60.6-67.1)	67.1 (65.5-68.7)
Rent housing (%)	36.2 (32.9-39.4)	32.7 (31.1-34.3)
Food stamps (%)	19.3 (17.1-21.5)	20 (19-21)
Households below poverty level (%)	22.3 (20.2-24.5)	23.5 (22.5-24.5)
Household median income (US$)	43,591 (40,090-47,092)	37,398.9 (36,210-38,588)
Families with one worker (%)	37.7 (36.4-39.1)	37.9 (37.1-38.6)
Families with two or more workers (%)	45.1 (43.1-47)	41.8 (40.8-42.7)
Urban area	0.36 (0.28-0.43)	0.16 (0.13-0.19)

[a] Other race includes Asian, Hawaiian, Pacific Islanders, and other races/ethnicities not defined as white, African American, or Hispanic.

Table 2. Factors associated with HIV hotspots, Mississippi, 2008-2014 (n=658).

Characteristic	Unadjusted model, OR (95% CI)	Adjusted model 1,[a] AOR[b] (95% CI)	Adjusted model 2,[c] AOR (95% CI)
Male (%)			
<48.25	Referent	Referent	Referent
≥48.25	0.70 (0.49-1.00)	0.94 (0.64-1.40)	0.94 (0.63-1.40)
Female (%)			
<51.70	Referent	Referent	Referent
≥51.70	1.43 (0.99-2.05)	—[d]	—
Total population			
<4199	Referent	Referent	Referent
≥4199	0.79 (0.55-1.13)	0.85 (0.55-1.31)	0.85 (0.55-1.32)
Median age (years)	0.96 (0.93-0.99)	0.96 (0.92-1.00)	0.95 (0.91-1.00)
<5	1.04 (0.96-1.12)	—	—
5-17	1.04 (0.99-1.08)	—	—
18-24	0.99 (0.97-1.02)	—	—
25-44	1.06 (1.02-1.10)	—	—
45-54	1.01 (0.95-1.07)	—	—
55-64	0.98 (0.93-1.04)	—	—
65-74	0.86 (0.80-0.92)	—	—
≥75	0.89 (0.82-0.96)	—	—
Less than a high school education (%)	0.96 (0.94-0.98)	0.95 (0.92-0.98)	0.91 (0.86-0.96)
Low ability to speak English (%)	1.03 (0.97-1.10)	—	—
White (%)			
<59.8	Referent	Referent	Referent
≥59.8	0.55 (0.38-0.79)	—	—
African American (%)			
<35.4	Referent	Referent	Referent
≥35.4	2.01 (1.39-2.90)	3.85 (2.23-6.65)	1.15 (0.42-3.17)
Hispanic (%)			
<1.4	Referent	Referent	Referent
≥1.4	0.77 (0.54-1.1)	0.86 (0.56-1.31)	0.84 (0.55-1.29)
Other ethnicity (%) [e]			
<0.6%	Referent	Referent	Referent
≥0.6%	1.11 (0.77-1.58)	—	—
Two or more race/ethnicities (%)			
<0.7	Referent	Referent	Referent
≥0.7	0.61 (0.43-0.88)	0.46 (0.30-0.70)	0.46 (0.30-0.72)
Population density			
<273	Referent	Referent	Referent
≥273	2.72 (1.87-3.97)	1.64 (0.96-2.80)	1.47 (0.86-2.51)
Median annual household income (US$)			
<36,775	Referent	Referent	Referent
≥36,775	1.26 (0.88-1.8)	1.12 (0.6-2.09)	0.80 (0.39-1.64)

Characteristic	Unadjusted model, OR (95% CI)	Adjusted model 1,[a] AOR[b] (95% CI)	Adjusted model 2,[c] AOR (95% CI)
Median annual individual income (US$)			
<20,083	Referent	Referent	Referent
≥20,083	1.53 (1.07-2.12)	—	—
Rent housing (%)	1.01 (1.0-1.019)	0.99 (0.97-1.01)	0.99 (0.97-1.01)
Own housing (%)	0.99 (0.98-1.00)	—	—
Households living at 100% below poverty status (%)			
<17.3	Referent	Referent	Referent
17.3-28.2	0.58 (0.37-0.90)	0.51 (0.28-0.92)	0.61 (0.33-1.13)
≥28.3	0.80 (0.52-1.22)	0.43 (0.19-0.99)	0.48 (0.20-1.17)
Households living below poverty level (%)	0.76 (0.53-1.08)	—	—
Urban area	2.89 (1.93-4.32)	2.01 (1.20-3.37)	2.06 (1.23-3.46)
≥ 2 workers in the family (%)			
<37.9	Referent	Referent	Referent
37.9-47.1	0.95 (0.59-1.52)		
≥47.2	1.78 (1.16-2.73)	—	—
1 worker in family (%)			
<34	Referent	Referent	Referent
34-40.6	0.79 (0.51-1.23)		
≥40.7	0.96 (0.62-1.47)	—	—
Households on food stamps (%)			
<17.76	Referent	Referent	Referent
≥17.76	0.88 (0.61-1.25)	—	—
Interaction between African American and less than a high school education	—	—	1.09 (1.02-1.16)

[a] Adjusted for percentage of the population living 100% below the poverty line, percentage of the total population that was male, median annual individual income, median annual household income, total population, percentage of total population that was African American, percentage of total population that was Hispanic, population density, households with two or more workers, median age, percentage of the total population that had less than a high school education, the percentage of the population that rented housing, and areas categorized as urban.

[b] AOR: adjusted odds ratio.

[c] Adjusted for the same as in adjusted model 1 and including the interaction variable between African American race and less than a high school graduate education.

[d] Indicates the variable was not statistically significant at a *P*<.25 level in bivariate analyses and therefore was not included in multivariable models.

[e] Other race includes Asian, Hawaiian, Pacific Islanders, and other races/ethnicities not defined as white, African American, or Hispanic.

The results of our descriptive statistical analyses are presented in Table 1. A total of 160 Mississippi census tracts were located within statistically significant HIV hotspot clusters, and 498 census tracts were located outside of HIV hotspot clusters.

In bivariate analyses focused on HIV rates for the entire Mississippi population, we noted significant associations between HIV hotspots and age, with positive associations between the percentage of the population that was 25 to 44 years of age (OR 1.06, 95% CI 1.02-1.10), census tracts with higher percentages of African American residents (OR 2.01, 95% CI 1.39-2.90), population density, median annual income, and residing in an urban area (Table 2). Significant protective associations were found between HIV hotspots and having less than a high school education (OR 0.96, 95% CI 0.94-0.98), being white (OR 0.55, 95% CI 0.38-0.79), identifying as two or more race/ethnicities (OR 0.61, 95% CI 0.43-0.88), and living below the poverty level (OR 0.58, 95% CI 0.37-0.90).

In the multivariable logistic regression models, we found that HIV hotspot clusters were significantly more likely to include census tracts that had a higher proportion of African Americans (adjusted odds ratio [AOR] 3.85, 95% CI 2.23-6.65) and urban census tracts (AOR 2.01, 95% CI 1.20-3.37), while controlling for all other factors. Hotspots were less likely to include census tracts with less than a high school education (AOR 0.96, 95% CI 0.94-0.98), people identifying as belonging to two or more racial/ethnic groups (AOR 0.46, 95% CI 0.30-0.70), and having

17.3% to 28.3% of the population living 100% below the poverty level (AOR 0.51, 95% CI 0.28-0.92).

In the second adjusted model, we included the interaction term for African American race and education, and found statistically significant associations with HIV hotspots and having less than a high school education (AOR 0.91, 95% CI 0.86-0.96), identifying as two or more races/ethnicities (AOR 0.46, 95% CI 0.30-0.72), and being located in an urban area (AOR 2.06, 95% CI 1.23-3.46). In this model, we noted a decrease in statistically significant associations between HIV hotspot clusters and other covariates. Thus, we included both multivariable models in Table 2 for comparison.

Discussion

Principal Findings

In this set of analyses conducted at the neighborhood (ie, census tract) level, we found that several geographic, demographic, and community-level factors were correlated with HIV hotspot clustering in Mississippi. The focus on clustering of all HIV infections, and infections among MSM, African Americans, and African American MSM allowed us to observe overlap and unique patterns in clustering across different populations. Both our multivariable models and geospatial hotspot cluster mapping suggests that in this largely rural state, urban residence is an important risk factor for HIV acquisition; most hotspot clusters—for African Americans, MSM, and African American MSM—were concentrated within or near the Jackson metropolitan area (Figures 5-7; outlined in green). This is consistent with the proposition that HIV spreads more rapidly in dense, highly populated sexual networks [17]. Indeed, research suggests that sexual networks may play a critical role in hastening HIV transmission, particularly among sexual and gender minorities in the United States and in Jackson, Mississippi [18-22].

In addition, we also found hotspot clusters in rural counties in the Mississippi Delta region, including Cahoma, Bolivar, and Sunflower Counties in the Mississippi Delta (see Figure 3). In contrast with the Jackson metropolitan area, hotspot clusters were not present in rural counties when analyses focused on MSM, suggesting that HIV transmission in these rural hotspots is attributed to heterosexual HIV transmission or different types of sexual networks. Another possibility is that there may be an underreporting of MSM contact in the Delta area, a rural area with high rates of HIV stigma. Lastly, we identified HIV hotspots among African Americans in Southern Mississippi, which may indicate higher transmission risks in the Gulf Coast Area, particularly among individuals who self-identify as heterosexual. We are uncertain whether there are strong network connections between the Gulf Coast Area and Jackson, or whether other mechanisms are in play in this region (eg, visitors from outside Mississippi who could bring elevated transmission risk). We believe that the sexual networks formed in urban and rural locations are associated with HIV transmission. Although sexual networks in urban locations may have increased densities compared to those in rural areas, we believe that urban-rural movement within sexual networks are associated with transmission patterns. A previous study in Mississippi documented this urban-rural pattern with phylogenetic analysis [23]. However, more research is needed given the lack of data on this rural phenomenon.

Taken together, these findings support the work of others that found that “place” is an important determinant of HIV acquisition risk [1,24-27] and that HIV is often concentrated in discrete geographic areas. Our analyses and results build on the descriptive maps presented in publicly available HIV data systems such as AIDSVu [28], which present HIV counts and rates at the county level. Our hotspot cluster analyses at the census tract level highlight statistically significant clusters of HIV at a more granular level (ie, census tract), which can be useful for targeting HIV prevention and care interventions in communities with high rates of HIV infection. Moreover, these findings underscore the need for different types of interventions in different geographic areas and among different subpopulations. As opposed to a static statewide intervention response, this cluster analysis allows for the strategic redirection of intervention efforts based on need rather than an assumption that all areas of the state warrant equal resources for HIV prevention. The cluster analysis methodology we employed could also be useful for other southern states.

Community-level factors were also associated with HIV hotspots. First, census tracts with between 17% and 28% of the population living below the federal poverty level had lower odds of being in HIV hotspot clusters, even after controlling for race, education, and other sociodemographic variables. This finding is congruent with other studies exploring concurrent sexual partnerships in Jackson, Mississippi, that found that higher income and higher education were associated with higher rates of overlapping sexual partnerships [29]. Similar trends have been found in Sub-Saharan Africa [30,31]. However, a large body of research in the United States found that HIV prevalence is higher among low-income individuals and communities, highlighting the differential role that income may play in distinct social contexts [32,33]. Although the mechanisms through which income and education may affect HIV acquisition risks have been studied previously [34], they are not yet well understood in the Deep South. Our findings suggest that higher income and education may be associated with increased number of sexual partnerships as well as geographic and social mobility, both of which can influence inner-connectivity of sexual networks. Low levels of income may also limit access to technology such as internet, mobile phones, and hookup apps, which can also expedite meeting of sexual partners and mobility within networks.

Most HIV prevention and care services in Mississippi are concentrated in Jackson. Further, most HIV screening that took place in 2015 in Mississippi was sponsored by the Mississippi Department of Health. Other parts of Mississippi have limited prevention and care services; far fewer services are delivered outside of Jackson. Larger sexual networks and limited access to HIV screening, treatment, and prevention services in urban geographic hotspots of infection likely compound HIV risks, particularly among MSM. Our findings underscore the need for expanding and tailoring HIV prevention and care services to the populations in urban hotspot clusters as well as less urbanized areas of the Mississippi Delta and the Gulf Coast

Region. Expanding the density of HIV testing locations to correspond with the identified hotspots is one important policy implication of the study findings.

These findings have important implications for implementation science and the public health response agenda focused on reducing racial and geographic disparities in HIV infection in the Deep South. Implementation science studies should explore how best to scale up existing HIV prevention, treatment, and care interventions in geographic hotspots in the Jackson metropolitan area, as well as in rural communities. Our results also suggest that culturally congruent interventions are likely important for different geographic regions: interventions in the Jackson metropolitan area should have a keen focus on engaging African American MSM in HIV prevention and care services, whereas interventions in the Delta may need greater focus on reaching heterosexual populations or reaching MSM who may not self-identify as gay.

A study of primary care physicians in the Mississippi Delta region suggests that expanding HIV screening to improve access is feasible [35]. Taken together, these findings underscore the urgency of engaging primary care providers in the Mississippi Delta effort to prevent HIV. Mississippi also has a robust effort to expand telemedicine, which could enhance access to HIV care for individuals in rural areas [36].

Limitations

Our study is subject to several limitations. First, a growing body of research suggests that complex sexual networks, particularly among African American MSM, contribute to HIV acquisition risks in Mississippi. However, surveillance data on HIV risk behaviors and sexual networks are not available and were therefore not included in our models. The surveillance data we used consisted of newly reported cases rather than incident cases; therefore, the HIV hotspots we identified do not necessarily reflect hotspots of recent HIV infection. Additionally, the location of HIV testing services may influence the location where cases are reported: fewer cases may be reported from an area located farther away from testing services because fewer people can access testing. We were able to geocode 80% of HIV cases in the Mississippi HIV surveillance data. Although this geocoding rate decreased the number of HIV cases that were included in our analyses, it is comparable to geocoding rates achieved in other studies focused on small area analysis in geospatial research [37]. Further, comparison of HIV cases that geocoded versus those that did not geocode were not significantly different providing further evidence that the data we analyzed are representative of the HIV epidemic in Mississippi between 2008 and 2014. Our results are based on ecological analyses. We did not have access to address-level data for the place of residence for HIV cases. Nonetheless, we were able to provide descriptive maps and cluster analysis results down to the census tract level, which can help to provide public health officials with a more detailed understanding of neighborhood-level geographic distribution of HIV clusters. Individual-level measures from the surveillance data were used to conduct spatial analyses focused on specific subgroups by race, sex, and age. Future research may benefit from multilevel hierarchical statistical models.

Future Directions

These study findings add to a mounting body of evidence underscoring how geographic location, community-level factors, and sexual orientation may impact HIV acquisition risks in the Deep South of the United States. We now have a list of important HIV prevention tools to deploy to reduce HIV transmission, including HIV testing, treatment as prevention, preexposure prophylaxis, and condom use. Our results could, for instance, help guide the locations where a mobile HIV testing clinic or a mobile clinic to prescribe preexposure prophylaxis could be positioned to address the local needs within hotspots. Our findings suggest a greater need for public health programs and implementation science research that explore how to effectively deploy these tools in “hotspot” communities in the Deep Southern United States with high rates of HIV infection and limited prevention, treatment, and care services.

Acknowledgments

This work was supported by grants P30AI42853, 1R34MH10937, K23AI096923, R25MH083620, R25DA035692, T32DA013911, and R25DA037190 from the National Institutes of Health.

Conflicts of Interest

None declared.

References

1. Nunn A, Yolken A, Cutler B, Trooskin S, Wilson P, Little S, et al. Geography should not be destiny: focusing HIV/AIDS implementation research and programs on microepidemics in US neighborhoods. Am J Public Health 2014 May;104(5):775-780 [FREE Full text] [doi: 10.2105/AJPH.2013.301864] [Medline: 24716570]
2. Centers for Disease Control and Prevention. 2015. Diagnoses of HIV infection in the United States and dependent areas URL: https://www.cdc.gov/hiv/library/reports/hiv-surveillance.html [accessed 2018-03-09] [WebCite Cache ID 6xnLh9ND9]
3. CDC Issue Brief: HIV in the Southern United States. Atlanta, GA: Centers for Disease Control and Prevention; 2016 May. URL: https://www.cdc.gov/hiv/pdf/policies/cdc-hiv-in-the-south-issue-brief.pdf [accessed 2018-03-09] [WebCite Cache ID 6xnLrv7g9]

4. Brouwer KC, Rusch ML, Weeks JR, Lozada R, Vera A, Magis-Rodríguez C, et al. Spatial epidemiology of HIV among injection drug users in Tijuana, Mexico. Ann Assoc Am Geogr 2012;102(5):1190-1199 [FREE Full text] [doi: 10.1080/00045608.2012.674896] [Medline: 23606753]
5. Stopka TJ, Lutnick A, Wenger LD, Deriemer K, Geraghty EM, Kral AH. Demographic, risk, and spatial factors associated with over-the-counter syringe purchase among injection drug users. Am J Epidemiol 2012 Jul 01;176(1):14-23 [FREE Full text] [doi: 10.1093/aje/kwr492] [Medline: 22562660]
6. Meyers DJ, Hood ME, Stopka TJ. HIV and hepatitis C mortality in Massachusetts, 2002-2011: spatial cluster and trend analysis of HIV and HCV using multiple cause of death. PLoS One 2014;9(12):e114822 [FREE Full text] [doi: 10.1371/journal.pone.0114822] [Medline: 25502820]
7. Centers for Disease Control and Prevention. 2016. HIV in the United States by geography URL: https://www.cdc.gov/hiv/statistics/overview/geographicdistribution.html [accessed 2018-03-09] [WebCite Cache ID 6xnM32r5F]
8. Centers for Disease Control and Prevention. 2017. NCHHSTP AtlasPlus URL: https://www.cdc.gov/nchhstp/atlas/index.htm [accessed 2018-03-09] [WebCite Cache ID 6xnMAKMto]
9. Centers for Disease Control and Prevention. 2015. HIV Surveillance reports URL: https://www.cdc.gov/hiv/library/reports/surveillance/ [accessed 2018-03-09] [WebCite Cache ID 6xnMJjpHt]
10. Mississippi State Department of Health. 2018. STD/HIV Office URL: https://msdh.ms.gov/msdhsite/_static/14,0,150.html [accessed 2018-03-09] [WebCite Cache ID 6xnMSX8NV]
11. Rosenberg ES, Grey JA, Sanchez TH, Sullivan PS. Rates of prevalent HIV infection, prevalent diagnoses, and new diagnoses among men who have sex with men in US states, metropolitan statistical areas, and counties, 2012-2013. JMIR Public Health Surveill 2016 May 17;2(1):e22 [FREE Full text] [doi: 10.2196/publichealth.5684] [Medline: 27244769]
12. Mena L, Crosby RA. Portrait of an epidemic: extremely high human immunodeficiency virus prevalence and incidence among young black men having sex with men and residing in a southern city. Sex Transm Dis 2017 Jul;44(7):401-402. [doi: 10.1097/OLQ.0000000000000606] [Medline: 28608789]
13. United States Census Bureau. Community facts URL: https://factfinder.census.gov/faces/nav/jsf/pages/index.xhtml [accessed 2018-03-09] [WebCite Cache ID 6xnMZ75SK]
14. Ord J, Getis A. Local spatial autocorrelation statistics: distributional issues and an application. Geogr Anal 1995 Oct;27(4):286-306 [FREE Full text] [doi: 10.1111/j.1538-4632.1995.tb00912.x]
15. Stopka TJ, Krawczyk C, Gradziel P, Geraghty EM. Use of spatial epidemiology and hot spot analysis to target women eligible for prenatal women, infants, and children services. Am J Public Health 2014 Feb;104 Suppl 1:S183-S189. [doi: 10.2105/AJPH.2013.301769] [Medline: 24354821]
16. Stopka TJ, Geraghty EM, Azari R, Gold EB, Deriemer K. Factors associated with presence of pharmacies and pharmacies that sell syringes over-the-counter in Los Angeles County. J Urban Health 2013 Dec;90(6):1079-1090 [FREE Full text] [doi: 10.1007/s11524-013-9798-7] [Medline: 23567984]
17. Rothenberg RB, Potterat JJ, Woodhouse DE, Muth SQ, Darrow WW, Klovdahl AS. Social network dynamics and HIV transmission. AIDS 1998 Aug 20;12(12):1529-1536. [Medline: 9727575]
18. Kenyon C, Colebunders R, Hens N. Determinants of generalized herpes simplex virus-2 epidemics: the role of sexual partner concurrency. Int J STD AIDS 2013 May;24(5):375-382. [doi: 10.1177/0956462412472816] [Medline: 23970705]
19. Omori R, Abu-Raddad L. Sexual network drivers of HIV and herpes simplex virus type 2 transmission. AIDS 2017 Jul 31;31(12):1721-1732 [FREE Full text] [doi: 10.1097/QAD.0000000000001542] [Medline: 28514276]
20. Morris M, Epstein H, Wawer M. Timing is everything: international variations in historical sexual partnership concurrency and HIV prevalence. PLoS One 2010 Nov 24;5(11):e14092 [FREE Full text] [doi: 10.1371/journal.pone.0014092] [Medline: 21124829]
21. Crosby RA, Mena L, Geter A. Favourable attitudes towards serosorting are associated with overall less frequent condom use among young Black men having sex men. Sex Health 2016 Feb;13(1):91-92. [doi: 10.1071/SH15156] [Medline: 26433647]
22. Nunn A, MacCarthy S, Barnett N, Rose J, Chan P, Yolken A, et al. Prevalence and predictors of concurrent sexual partnerships in a predominantly African American population in Jackson, Mississippi. AIDS Behav 2014 Dec;18(12):2457-2468 [FREE Full text] [doi: 10.1007/s10461-014-0777-0] [Medline: 24803130]
23. Oster AM, Pieniazek D, Zhang X, Switzer WM, Ziebell RA, Mena LA, et al. Demographic but not geographic insularity in HIV transmission among young black MSM. AIDS 2011 Nov 13;25(17):2157-2165. [doi: 10.1097/QAD.0b013e32834bfde9] [Medline: 21866038]
24. LaVeist T, Pollack K, Thorpe R, Fesahazion R, Gaskin D. Place, not race: disparities dissipate in southwest Baltimore when blacks and whites live under similar conditions. Health Aff (Millwood) 2011 Oct;30(10):1880-1887. [doi: 10.1377/hlthaff.2011.0640] [Medline: 21976330]
25. Vaughan AS, Rosenberg E, Shouse RL, Sullivan PS. Connecting race and place: a county-level analysis of White, Black, and Hispanic HIV prevalence, poverty, and level of urbanization. Am J Public Health 2014 Jul;104(7):e77-e84 [FREE Full text] [doi: 10.2105/AJPH.2014.301997] [Medline: 24832420]

26. Nunn A, Towey C, Chan PA, Parker S, Nichols E, Oleskey P, et al. Routine HIV screening in an urban community health center: results from a geographically focused implementation science program. Public Health Rep 2016;131 Suppl 1:30-40 [FREE Full text] [doi: 10.1177/00333549161310S105] [Medline: 26862228]
27. Chan PA, Towey C, Poceta J, Rose J, Bertrand T, Kantor R, et al. Online hookup sites for meeting sexual partners among men who have sex with men in Rhode Island, 2013: a call for public health action. Public Health Rep 2016;131(2):264-271 [FREE Full text] [doi: 10.1177/003335491613100210] [Medline: 26957661]
28. AIDSVu. URL: https://aidsvu.org/ [accessed 2018-03-19] [WebCite Cache ID 6y2MLHcXe]
29. Marshall BD, Perez-Brumer AG, MacCarthy S, Mena L, Chan PA, Towey C, et al. Individual and partner-level factors associated with condom non-use among African American STI clinic attendees in the Deep South: an event-level analysis. AIDS Behav 2016 Jun;20(6):1334-1342 [FREE Full text] [doi: 10.1007/s10461-015-1266-9] [Medline: 26683032]
30. Rodrigo C, Rajapakse S. HIV, poverty and women. Int Health 2010 Mar;2(1):9-16. [doi: 10.1016/j.inhe.2009.12.003] [Medline: 24037044]
31. Gillespie S, Kadiyala S, Greener R. Is poverty or wealth driving HIV transmission? AIDS 2007 Nov;21 Suppl 7:S5-S16. [doi: 10.1097/01.aids.0000300531.74730.72] [Medline: 18040165]
32. Karon JM, Fleming PL, Steketee RW, DeCock KM. HIV in the United States at the turn of the century: an epidemic in transition. Am J Public Health 2001 Jul;91(7):1060-1068. [Medline: 11441732]
33. Adimora AA, Schoenbach VJ, Doherty IA. HIV and African Americans in the southern United States: sexual networks and social context. Sex Transm Dis 2006 Jul;33(7 Suppl):S39-S45. [doi: 10.1097/01.olq.0000228298.07826.68] [Medline: 16794554]
34. Zierler S, Krieger N, Tang Y, Coady W, Siegfried E, DeMaria A, et al. Economic deprivation and AIDS incidence in Massachusetts. Am J Public Health 2000 Jul;90(7):1064-1073. [Medline: 10897184]
35. Sison N, Yolken A, Poceta J, Mena L, Chan PA, Barnes A, et al. Healthcare provider attitudes, practices, and recommendations for enhancing routine HIV testing and linkage to care in the Mississippi Delta region. AIDS Patient Care STDS 2013 Sep;27(9):511-517 [FREE Full text] [doi: 10.1089/apc.2013.0169] [Medline: 23991689]
36. Kempf M, Huang C, Savage R, Safren SA. Technology-delivered mental health interventions for people living with HIV/AIDS (PLWHA): a review of recent advances. Curr HIV/AIDS Rep 2015 Dec;12(4):472-480 [FREE Full text] [doi: 10.1007/s11904-015-0292-6] [Medline: 26452644]
37. Krieger N, Chen JT, Waterman PD, Rehkopf DH, Subramanian SV. Race/ethnicity, gender, and monitoring socioeconomic gradients in health: a comparison of area-based socioeconomic measures--the public health disparities geocoding project. Am J Public Health 2003 Oct;93(10):1655-1671. [Medline: 14534218]

Abbreviations

ACS: American Community Survey
AOR: adjusted odds ratio
GIS: geographic information system
MSM: men who have sex with men

Edited by T Sanchez; submitted 04.09.17; peer-reviewed by R Valdiserri, K Card; comments to author 10.10.17; revised version received 21.01.18; accepted 04.02.18; published 03.04.18.

Please cite as:
Stopka TJ, Brinkley-Rubinstein L, Johnson K, Chan PA, Hutcheson M, Crosby R, Burke D, Mena L, Nunn A
HIV Clustering in Mississippi: Spatial Epidemiological Study to Inform Implementation Science in the Deep South
JMIR Public Health Surveill 2018;4(2):e35
URL: http://publichealth.jmir.org/2018/2/e35/
doi:10.2196/publichealth.8773
PMID:29615383

Original Paper

Predictors of Facebook User Engagement With Health-Related Content for Gay, Bisexual, and Other Men Who Have Sex With Men: Content Analysis

Kiffer George Card[1,2], PhD; Nathan Lachowsky[2,3], PhD; Blake W Hawkins[4], MLIS; Jody Jollimore[5], MPP; Fahmy Baharuddin[6]; Robert S Hogg[1,2], PhD

[1]Faculty of Health Science, Simon Fraser University, Burnaby, BC, Canada
[2]British Columbia Centre for Excellence in HIV/AIDS, Vancouver, BC, Canada
[3]School of Public Health and Social Policy, University of Victoria, Victoria, BC, Canada
[4]Interdisciplinary Studies Graduate Program, University of British Columbia, Vancouver, BC, Canada
[5]Community-Based Research Centre for Gay Men's Health, Vancouver, BC, Canada
[6]YouthCO HIV & Hep C Society, Vancouver, BC, Canada

Corresponding Author:
Kiffer George Card, PhD
Faculty of Health Science
Simon Fraser University
c/o Faculty of Health Sciences
8888 University Drive
Burnaby, BC, V5A 1S6
Canada
Phone: 1 7789884269
Email: kiffercard@gmail.com

Abstract

Background: Social media is used by community-based organizations (CBOs) to promote the well-being of gay and bisexual men (GBM). However, few studies have quantified which factors facilitate the diffusion of health content tailored for sexual minorities.

Objective: The aim of this study was to identify post characteristics that can be leveraged to optimize the health promotion efforts of CBOs on Facebook.

Methods: The Facebook application programming interface was used to collect 5 years’ of posts shared across 10 Facebook pages administered by Vancouver-based CBOs promoting GBM health. Network analysis assessed basic indicators of network structure. Content analyses were conducted using informatics-based approaches. Hierarchical negative binomial regression of post engagement data was used to identify meaningful covariates of engagement.

Results: In total, 14,071 posts were shared and 21,537 users engaged with these posts. Most users (n=13,315) engaged only once. There was moderate correlation between the number of posts and the number of CBOs users engaged with (r=.53, P<.001). Higher user engagement was positively associated with positive sentiment, sharing multimedia, and posting about pre-exposure prophylaxis, stigma, and mental health. Engagement was negatively associated with asking questions, posting about dating, and sharing posts during or after work (versus before).

Conclusions: Results highlight the existence of a core group of Facebook users who facilitate diffusion. Factors associated with greater user engagement present CBOs with a number of strategies for improving the diffusion of health content.

(JMIR Public Health Surveill 2018;4(2):e38) doi:10.2196/publichealth.8145

KEYWORDS

social media; health promotion; gay and bisexual men; user engagement

Introduction

Gay, bisexual, and other men who have sex with men (GBM) are at elevated risk for a number of adverse health outcomes [1,2]. Stall et al [3] argues that gay communities experience a syndemic of co-occurring sexual, substance use, and psychosocial conditions that, according to Singer [4], work synergistically under “deleterious social and physical conditions” (p 15) to adversely affect the health of this population [5]. In response, public health and community leaders have advanced holistic approaches to gay men’s health that address not only individual and biological factors, but also the broader psychosocial and structural factors that affect their health and well-being [6].

In implementing such programs, social media is widely used by community-based organizations (CBOs) to disseminate health information and engage with GBM [7-9]. Indeed, social media has come to play a significant and diverse role in a variety of health contexts. Articulating this role, Kietzmann et al [10] highlight seven personal and interpersonal needs that social media has come to fulfill. Broadly, we summarize these needs by three activities: identity management, communication, and social bonding. In the context of GBM health, sexual minorities have always needed spaces where they can engage in these activities, and social media has come to provide such spaces [11,12].

Although the Internet provides a platform whereby CBOs can reach GBM, the success of these interventions is far from guaranteed [13]. Rogers’ Diffusion of Innovations Theory describes the challenges to CBOs in terms of diffusion, reach, and uptake [14,15]. In brief, Rogers posits that key characteristics of individuals (whom he describes as “adopters”) and the network ties that connect them to others in a social network are fundamental to the spread of information, behavior, and products. A number of factors have been identified that impact adoption and diffusion (eg, age, social network structure, personality types), and media richness theory describes how specific media (ie, routes of content delivery) detract or promote diffusion [16]. Furthermore, he argues that more “life-like” interactions better promote uptake of new ideas.

In the age of social media, specific engagement indicators (ie, reactions, comments, and shares) on Facebook provide rudimentary markers for diffusion—and, in fact, are used by Facebook’s EdgeRank algorithm to govern which messages are shown to other users [17]. Barriers to diffusion are particularly relevant to efforts targeting GBM, who represent a diverse and uniquely organized group of individuals [18]. For example, Cassidy [19] notes that campaigns to amass likes, comments, and shares can often be at odds with an individual’s need to manage their public identity. After all, not all sexual minorities openly acknowledge their sexuality online—especially in spaces where multiple social circles collide [20]. Yet, if social media strategies among GBM are to be successful, CBOs must find ways to encourage users to engage with their content. This is because many social media platforms rely on engagement-based algorithms to determine if social media content is viewed by other users. For example, according to Facebook:

> *The stories that show in your News Feed are influenced by your connections and activity on Facebook. This helps you to see more stories that interest you from friends you interact with the most. The number of comments and likes a post receives and what kind of story it is (ex: photo, video, status update) can also make it more likely to appear in your News Feed.* [17]

Consistent with this, increasing user engagement (defined by Facebook as the composite of reactions, comments, and shares on a post) has become a primary objective of social media campaigns, and a handful of studies have sought to identify predictors of user engagement. For example, Veale et al [21] identified 10 Twitter and Facebook profiles with high user engagement and found that these organizations gained prominence by posting regularly, engaging with individual users, encouraging interaction and conversation by posing questions, sharing multimedia, and highlighting celebrity involvement. In a similar study, Kite et al [22] found that higher post engagement among 20 Facebook health profiles was associated with positive sentiment, providing factual information, inclusion of videos, and celebrity marketing. Likewise, Rus and Cameron [23] explored 10 diabetes-related health pages and found that imagery was a strong predictor of engagement. Further, they identified other characteristics, such as sentiment, crowdsourcing, and providing factual information, that were associated with some, but not all, forms of engagement. However, as campaigns addressing sensitive subjects and those targeting sexual minorities might be uniquely constrained by users’ willingness to publicly endorse or share CBO-generated content, context-specific evaluations of user engagement are needed. As such, the primary objective of this study was to identify strategies to enhance user engagement.

Additionally, it is unclear whether Facebook is even an effective platform for CBOs to reach sexual minority populations [24]. Indeed, although social media campaigns might gain the attention of local network members, they may miss those who are not directly associated with CBOs. Despite widely held assumptions of Facebook’s communication potential [9], little research has been conducted on the Facebook network structure of sexual minorities. Optimistically, that which has suggests that the Facebook network structure of sexual minorities is scale-free [25], meaning that some individuals are more embedded in the social network than others and that these individuals act like “hubs” diffusing information into their local networks. However, although scale-free networks are said to effectively transmit information [26], their efficiency relies on the ways these networks are organized [14]. For example, scale-free networks with high modularity (ie, the appearance of distinct clusters or communities within a network) promote strong bonds between network members and thus saturation of local networks, whereas those with low modularity promote weak ties between individuals, but broad global diffusion [27,28]. Both modular and nonmodular network structures offer benefits and limitations; for example, experimental research by Bakshy et al [27] shows that strong ties increase the likelihood that individuals will share content shared by other network members, whereas weak ties facilitate the diffusion of information between network clusters. Therefore, as a secondary objective, this study aimed to complement our understanding of the diffusion of information through the Facebook networks of CBOs in Vancouver, British Columbia (BC).

Methods

Consistent with these objectives, this study leveraged data collected from 10 Facebook pages (ie, all pages identified as being administered by selected organizations) belonging to eight CBOs in Vancouver, BC. Pages were purposively selected (ie, all identified organizations were included) that were (1) well known to our study team (ie, community-based partners or those otherwise highly visible), (2) inclusive of or targeted toward sexual minorities (ie, page content relevant, at least in part, to sexuality, sexual health, or community social issues), and (3) dealt primarily with health promotion (ie, health promotion was main goal of the organization).

To ensure user privacy and compliance with Facebook's end-user agreement, data were downloaded using Facebook's public application programming interface (API) accessed through the Netvizz Facebook app [29]. Data collected between January 1, 2010 and August 31, 2016, via Netvizz were hierarchically organized by page and post. The first year—2010—was selected based on the completion of the iPrEx trial examining the efficacy of pre-exposure prophylaxis (PrEP), one of the key topics assessed in this analysis [30]. At the page level, we identified the number of followers for each page. On the post level, we identified the number of likes, comments, and shares on each post. Netvizz also assigned unique identifiers to each user, allowing us to examine user engagement across multiple posts and multiple pages. As such, we used Spearman rank correlation to determine whether there was an association between frequency of participation and participation across multiple pages. Further, a network diagram showing the ways individuals interacted with posts from the 10 CBOs was constructed in Gephi 0.9.1 using the ForceAtlas2 layout algorithm [31]. Modularity clusters were also identified using Gephi's modularity tool with the resolution set to 1 in order to maximize the modularity [32]. Because this study leveraged publicly available data, the research ethics board at Simon Fraser University deemed the study exempt from review. As an extra precaution on behalf of the users whose data were included in the present analysis, the names of the Facebook pages included in our study have been omitted. However, Table 1 provides a short characterization of the mission of each page to highlight the range of groups included in our analysis.

The content of each post was then analyzed using informatics-based methodology [33-35]. First, using researcher-generated search taxonomies, we identified posts relating to eight topics (with keywords for each topic in parentheses): pre-exposure prophylaxis (ie, PrEP, preexposure, pre-exposure, prophylaxis), treatment (ie, treatment, undetect*, viral load, viral-load), condoms (ie, condom*), mental health (ie, mental, emotion*, depress*, anxiety), stigma (ie, stigma, discriminat*), testing (ie, test*, screening, checked online), dating (ie, dating, relationship), and research (ie, research*, study). Posts that utilized questions to engage users were also recorded by identifying posts with a question mark (ie, "?"). Similarly, posts which directly encouraged user engagement were identified by searching for key terms inviting participation (ie, like, comment, share, take, visit).

Further, each sentence of each post was scored using the Bing Liu sentiment lexicon [34]. The Bing Liu sentiment lexicon, which is widely used in sentiment analysis and opinion mining, was selected because it provides a freely accessible word database that assigns positive and negative values to keywords, including commonly misspelled words. After each word within each sentence was scored, an average sentiment score was assigned to each post indicating whether the post had an overall negative or positive affect.

We then used multivariable hierarchical negative binomial regression to identify the post characteristics associated with greater user engagement. In this analysis, Facebook's engagement score was used because this is presumably an important variable used in their News Feed algorithm. According to Facebook's API, the number is calculated as the combined total number of reactions, shares, and comments on each post. Hierarchical negative binomial regression modeling was selected as the statistical approach for this study because the Facebook engagement count data were overdispersed, highly skewed toward 0 and 1, and came from 10 separate Facebook pages—each with a varying number of Facebook "fans" and with differing rates of activity. Incidence rate ratios (IRR) presented in text were calculated by exponentiating the regression coefficients. All coding and statistical analysis were conducted in RStudio.

Results

Table 1 provides a basic description for each of the 10 Facebook pages included in our study, including the number of posts shared by each organization. Table 2 provides an overview of the posts analyzed in this study. During the study period between January 1, 2010 and August 31, 2016, 14,071 posts were shared. In total, 21,537 unique users were identified as having engaged with at least one post. Most users engaged only once (n=13,315), two to five times (n=4872), or six to nine times (n=1197). Approximately 10% (2153/21,537) of users engaged more than 10 times.

Similarly, most users engaged with content from only one (n=18,837) or two (n=1978) groups. Only a small minority of users (n=722) interacted with more than three groups. Despite low overall engagement (low number of users who "engaged" with content more than once), high modularity (Q=0.62) was observed in the ways individuals interacted with shared content (see Figure 1). Indeed, eight modularity clusters accounted for 74.49% of posts (10,481/14,071) and 93.31% of users (20,097/21,537). There was moderate correlation between the number of posts and the number of CBOs users engaged with (r=.53, P<.001).

Table 1. Description of selected Facebook pages serving Vancouver's gay communities. LGBT2SQ: lesbian, gay, bisexual, transgender, two-spirited, queer. IQR: interquartile range.

ID and description of organization		Months of observation[a]	"Facebook fans"	Posts shared by organization	Engagements
1	AIDS service organization	78	1168	1458	1281
2a	Gay men's research organization	47	422	578	262
2b	Anti-stigma social media campaign[b]	10	1126	242	524
3	AIDS fundraising organization	79	1578	699	1385
4	Gay men's health organization	80	2290	2166	3184
5	LGBT2SQ Pride organization	74	8813	1791	7405
6	AIDS service organization	65	1015	1921	675
7	Queer community organization	80	5571	3607	9351
8a	Youth-led health organization	74	1097	758	824
8b	Peer-led program for young LGBT2SQ[b]	54	716	851	598
Median (IQR)		74 (57-79)	1147 (1036-2112)	1155 (714-1889)	1053 (617-2734)
Total		641	23,796	14,071	25,489

[a] Months of observation indicate the total number of months the page was operational for, with organizations new to the Facebook platform providing fewer months of observation.

[b] These pages are associated with the Facebook page listed before (ie, are administered by these groups as subprograms, but for marketing reasons are separate from the main page administered by the organization).

Table 2. Post characteristics across 10 Facebook pages serving Vancouver's gay communities, 2010-2016. IQR: interquartile range.

Post characteristics	Facebook posts
Post type, n (%)	
Status	1805 (12.83)
Photo	3280 (23.31)
Video	481 (3.42)
Link	7666 (54.48)
Event	839 (5.96)
Health message, n (%)	
Pre-exposure prophylaxis	119 (0.80)
Stigma	305 (2.20)
Mental health	180 (1.30)
Treatment	215 (1.50)
Testing	403 (2.90)
Research	380 (2.70)
Condoms	241 (1.70)
Dating	253 (1.80)
Time of week, n (%)	
Weekday (Monday-Friday)	12,368 (87.90)
Weekend (Saturday and Sunday)	1703 (12.10)
Time of day, n (%)	
Before work (1:00 am-7:59 am)	241 (1.71)
During work (8:00 am-4:59 pm)	10,734 (76.28)
After work (5:00 pm-12:59 am)	3096 (22.00)
Post feature	
Sentiment, median (IQR[a])	0.09 (0.00-0.27)
Questions, n (%)	2824 (20.10)
Direct invitations to participate, n (%)	2326 (16.50)

Figure 1. Network diagram illustrating user engagement with each post. Colors represent modularity clusters. Numbered symbols represent each Facebook page with the location indicating the modularity class in which most posts were located.

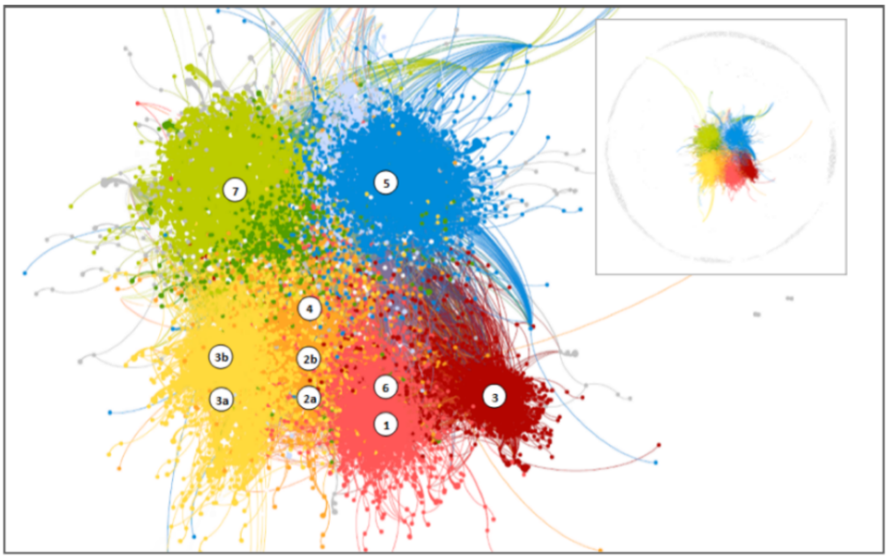
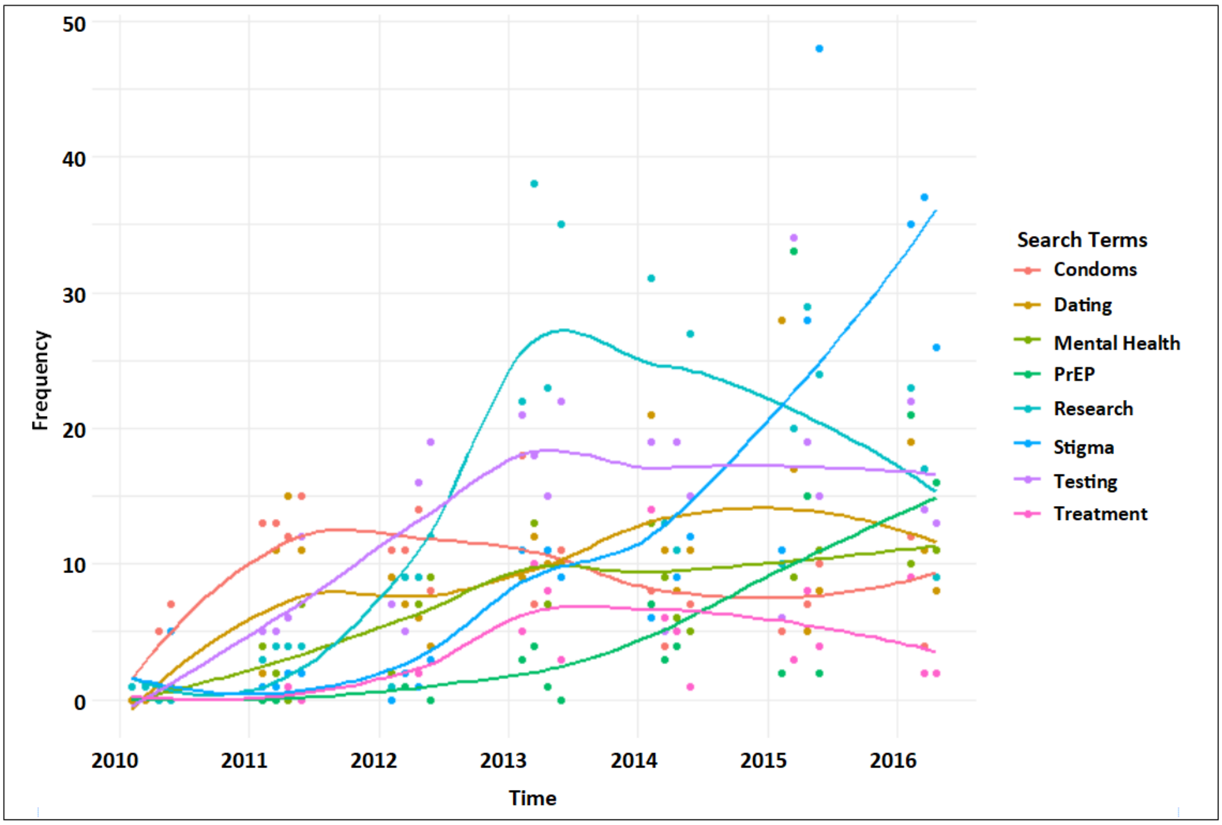


Table 3. Factors associated with user engagement.

Predictors of “engagement”	Regression coefficient B (SE)	Incidence rate ratio	*P* value
Post type			
Status	Reference		
Photo	1.10 (0.04)	3.00	<.001
Video	0.84 (0.07)	2.32	<.001
Link	0.51 (0.04)	1.66	<.001
Event	–0.36 (0.06)	0.70	<.0001
Health message (yes vs no)			
Pre-exposure prophylaxis	1.29 (0.12)	3.64	<.001
Stigma	0.47 (0.08)	1.60	<.001
Mental health	0.42 (0.11)	1.52	<.001
Treatment	0.16 (0.10)	1.17	.10
Testing	0.14 (0.07)	1.15	.06
Research	0.01 (0.08)	1.01	.90
Condoms	–0.06 (0.10)	0.94	.55
Dating	–0.33 (0.09)	0.72	<.001
Time of week			
Weekday (Monday-Friday)	Reference		
Weekend (Saturday and Sunday)	0.07 (0.04)	1.07	.049
Time of day			
Before work (1:00 am-7:59 am)	Reference		
During work (8:00 am-4:59 pm)	–0.27 (0.09)	0.76	<.001
After work (5:00 pm-11:59 pm)	–0.23 (0.09)	0.79	.01
Post feature (yes vs no)			
Sentiment	0.52 (0.05)	1.68	<.001
Questions	–0.10 (0.03)	0.90	<.001
Direct invitations to participate	–0.06 (0.03)	0.94	.07

Post characteristics associated with user engagement are shown in Table 3. Higher user engagement was positively associated with positive sentiment (IRR 1.68), sharing photos (IRR 3.00), videos (IRR 2.32), and links (IRR 1.66), and posting about PrEP (IRR 3.64), stigma (IRR 1.60), and mental health (IRR 1.52). Figure 2 shows the frequency of health messaging over time for the key terms assessed in this analysis. Engagement was negatively associated with asking a question (IRR 0.90), posting about dating (IRR 0.72), sharing posts during (IRR 0.76) or after work (IRR 0.79) compared to before work and with sharing events (IRR 0.70).

Figure 2. Loess smoothed mention of health messages overtime (2010-2016), stratified by keyword. PrEP: pre-exposure prophylaxis.



Discussion

This study collected post data from 10 Facebook pages promoting health or health-related events to GBM in Vancouver, BC. Together, these 10 pages had approximately 24,000 followers, shared approximately 14,000 posts, and amassed more than 25,000 engagements (ie, likes, comments, shares) during the 7 years' of data analyzed. Although our data do not speak empirically to the true network structure of Facebook's gay communities in Vancouver, we can make several important inferences regarding the network structure that underlies this analysis. First, based on the correlation between the number of groups and the number of engagements, our results point to the existence of a core group of users who may promote the diffusion of health content. Indeed, only a minority (38.2%) of users engaged more than once over the 7-year period we studied. These observations suggest that the true Facebook network structure of Vancouver's gay community is indeed scale-free, as shown by Silenzio et al [25]. Second, because most users only engaged once over the extended timeframe of this analysis, our findings also suggest that shared content is broadly diffusing into distal regions of the network among individuals who may not be directly linked to the Facebook pages included in this analysis [28]. Third, as we observed modularity in user-post engagement, our findings also suggest that the 10 Facebook pages included in this analysis are serving multiple, distinct, although linked, clusters. Indeed, although some Facebook pages overlap in their outreach, our findings (see Figure 1) suggest that the combined effort of these organizations reaches into distinct user communities. This suggests that both strong and weak ties make the Facebook platform an ideal location for the diffusion of health content [27].

Our analysis also identified several factors that may enhance the diffusion of health content by increasing user engagement. These findings may be of help to CBOs because, unlike social network factors, they are amenable to intervention and change. For instance, we found that posts shared in the morning diffused better than those shared during working hours or after work. These results are consistent with previous studies that showed that posts can be strategically timed to take advantage of when users are active. Similarly, the richness of posts was also shown to be an important covariate of user engagement with higher engagement associated with photos, videos, and links, and lower engagement associated with sharing events. This is consistent with previous research [21] and with media richness theory [36], which suggests that "richer" media (ie, those with greater ability to efficiently convey messages, social cues, personalization, and feedback) better engages target audiences.

However, contradicting this theory, we also found that specific strategies to engage users, such as asking questions, were associated with lower user engagement. This supports other research that shows that inviting engagement, ironically, may be a less effective way to promote engagement [23]. Other research has shown more generally that traditional marketing elements discourage user engagement on Facebook [22]. This may reflect a distrust for traditional marketing and a desire for more authentic communication [37]. Indeed, Fromm et al [38] recommend that marketers approach younger audiences not as target populations, but as partners in the advertising process. Consistent with this approach, social media strategies should identify ways to authentically promote health with, not to, GBM [21]. Posts with positive affect did elicit higher engagement—perhaps reflecting the well-documented heuristic bias toward positive messaging [23,39].

Closely related to the form of posts, the content of posts was also seen to have a significant effect on user engagement. Posts about PrEP, stigma, and mental health exhibited greater engagement, whereas posts about dating had lower engagement. Although it is difficult to assess why some subjects engaged users better in this research, these finding may reflect the health priorities, or perhaps current controversies, in gay communities. Therefore, higher user engagement is expected when pages are posting content that might be trending and amenable to gay communities—highlighting the importance of community-conscious agendas for health promotion. Indeed, during the time of this study, community-driven campaigns around PrEP [40] and stigma [41] may have served as driving forces behind user engagement with posts regarding PrEP, stigma, and mental health. Conversely, posts relating content regarding HIV-related behaviors (eg, testing and condoms) seemed to attract fewer engagements, potentially highlighting the difficulty of using social media to promote well-established prevention strategies. This may be particularly true for those with which audiences have become fatigued, such as has long been reported among GBM in San Francisco [42]. Based on our results, future analyses should investigate whether integrating better diffusing content, such as PrEP and stigma, into posts promoting more traditional prevention strategies has the potential to improve the diffusion of this content.

Regular assessment of how users are engaging with posts relevant to specific key themes may provide public health and community leaders with insight into the diffusion of social discourse surrounding important topics of concern. To this point, we note significant temporal variation in the frequency at which key themes were included in CBO posts. As mentioned before, PrEP and stigma increased throughout the observation period likely due to specific prevention campaigns in Metro Vancouver. Similarly, the frequency at which research and testing were discussed increased dramatically during the first half of the observation period, with research-related posts peaking in early 2013 and declining thereafter, and testing-related posts leveling off at the same time. Because this study was primarily focused on engagement and not the CBO's rationale for content selection, future studies might improve our understanding of what factors contribute to the ebb and flow of specific key themes.

Further, future research should examine individual-level data, particularly that of core users, whom our findings suggest may play an important role in the diffusion of post content. Such examinations might be conducted by each CBO because they may have greater access and interest in these specific analyses. More generally, our findings also highlight the importance of the user experience in shaping the diffusion of health content. Therefore, ongoing cooperation with users is needed to identify the features that should be leveraged in health promotion—especially because users, not social media specialists, are the ultimate arbiters of whether content is shared with their networks. Consistent with this, CBOs may benefit from examining the network dynamics of their followers and leverage the approaches used in this study to identify specific users who might be willing to partner with CBOs to promote their content.

These findings should be interpreted with consideration of the limitations of this study. First, because CBOs were not selected using a randomized approach, it is difficult to say whether our findings are generalizable to all Facebook-based health promotion efforts. However, we included most of the major pages associated with organizations in Metro Vancouver. Therefore, our results best represent the health priorities of Vancouver's gay community, although they may not be the same as those in other communities. Second, because we used relatively simple informatics-based analytic approaches to identify and code posts, our analysis is subject to measurement error. In particular, the selection of key terms may limit the accurate classification of posts relevant to the post features and health messages we explored. However, based on the consistency of our findings with studies conducted regarding other health areas, it seems that our approach produced similar results to studies that included manual coding techniques [22,23]. Nevertheless, validation of the results of this study is needed, both in other geographic settings and with other sexual or gender minority communities. Third, because the engagement factors for Facebook reactions, comments, and shares may differ [21,23], further analysis is needed on how to elicit the type of participation that will best promote health awareness. This is especially important given that the predictors of likes, comments, and shares may not be the same. Indeed, because we summed across these three types of user engagement, we may be obscuring important differences or patterns. For example, posts that elicit comments may elicit fewer shares, thus misestimating user engagement with shared posts. Furthermore, Facebook's EdgeRank algorithm, which determines whether content is diffused and shown on people's Facebook pages, is constantly updated and the relative weighting of various types of interaction may change, making it important to understand the unique determinants of various types of engagement (ie, reactions, comments, shares). Future analyses should expand our findings by evaluating the factors associated with specific engagement indicators. Lastly, other important factors, which we have not considered, may also shape user engagement. These include individual-level factors, which require a different analytic and sampling approach to understand how specific user characteristics may shape user engagement. Although engagement at the individual level is difficult to study, integrating Facebook plug-ins into study questionnaires might allow researchers to match social media participation to survey responses. Other important considerations may also include specific factors that might persuade different individuals to engage with post content, underscoring the need for further examination of gay and bisexual men's social media engagement. Likewise, exploration of additional themes that were not examined in this analysis is needed. Indeed, only a minority of posts were relevant to the themes we selected and examined. Undoubtedly, CBOs have interest in sharing and promoting content that may not necessarily be directly related to health outcomes studied by public health researchers. Despite these limitations, this study supports the use of Facebook for health promotion among sexual minorities and highlights multiple factors that can be leveraged to optimize user engagement, thus enhancing the diffusion of health information and the reach of CBOs.

Acknowledgments

KGC is supported by the Momentum Health Study as part of his doctoral training. Momentum is funded through the National Institute on Drug Abuse (R01DA031055-01A1) and the Canadian Institutes for Health Research (MOP-107544, 143342).

Conflicts of Interest

None declared.

References

1. Isacco A, Yallum NK, Chromik LC. A review of gay men’s health. AM J Lifestyle Med 2011 Dec 12;6(1):45-62. [doi: 10.1177/1559827611402580]
2. Lee R. Health care problems of lesbian, gay, bisexual, and transgender patients. West J Med 2000 Jun;172(6):403-408 [FREE Full text] [Medline: 10854396]
3. Stall R, Mills TC, Williamson J, Hart T, Greenwood G, Paul J, et al. Association of co-occurring psychosocial health problems and increased vulnerability to HIV/AIDS among urban men who have sex with men. Am J Public Health 2003 Jun;93(6):939-942. [Medline: 12773359]
4. Singer M. Pathogen-pathogen interaction: a syndemic model of complex biosocial processes in disease. Virulence 2010;1(1):10-18 [FREE Full text] [doi: 10.4161/viru.1.1.9933] [Medline: 21178409]
5. Singer M, Bulled N, Ostrach B, Mendenhall E. Syndemics and the biosocial conception of health. Lancet 2017 Mar 04;389(10072):941-950. [doi: 10.1016/S0140-6736(17)30003-X] [Medline: 28271845]
6. Halkitis PN, Wolitski RJ, Millett GA. A holistic approach to addressing HIV infection disparities in gay, bisexual, and other men who have sex with men. Am Psychol 2013;68(4):261-273. [doi: 10.1037/a0032746] [Medline: 23688093]
7. Noar SM, Willoughby JF. eHealth interventions for HIV prevention. AIDS Care 2012;24(8):945-952. [doi: 10.1080/09540121.2012.668167] [Medline: 22519523]
8. Schnall R, Travers J, Rojas M, Carballo-Diéguez A. eHealth interventions for HIV prevention in high-risk men who have sex with men: a systematic review. J Med Internet Res 2014;16(5):e134 [FREE Full text] [doi: 10.2196/jmir.3393] [Medline: 24862459]
9. Taggart T, Grewe M, Conserve D, Gliwa C, Roman IM. Social media and HIV: a systematic review of uses of social media in HIV communication. J Med Internet Res 2015 Nov 02;17(11):e248 [FREE Full text] [doi: 10.2196/jmir.4387] [Medline: 26525289]
10. Kietzmann JH, Hermkens K, McCarthy IP, Silvestre BS. Social media? Get serious! Understanding the functional building blocks of social media. Bus Horizons 2011 May;54(3):241-251. [doi: 10.1016/j.bushor.2011.01.005]
11. Grov C, Breslow AS, Newcomb ME, Rosenberger JG, Bauermeister JA. Gay and bisexual men's use of the Internet: research from the 1990s through 2013. J Sex Res 2014;51(4):390-409 [FREE Full text] [doi: 10.1080/00224499.2013.871626] [Medline: 24754360]
12. Harper GW, Serrano PA, Bruce D, Bauermeister JA. The Internet's multiple roles in facilitating the sexual orientation identity development of gay and bisexual male adolescents. Am J Mens Health 2016 Sep;10(5):359-376 [FREE Full text] [doi: 10.1177/1557988314566227] [Medline: 25585861]
13. Korda H, Itani Z. Harnessing social media for health promotion and behavior change. Health Promot Pract 2013 Jan;14(1):15-23. [doi: 10.1177/1524839911405850] [Medline: 21558472]
14. Rogers M, Chapman C, Giotsas V. Measuring the diffusion of marketing messages across a social network. J Direct Data Digit Mark Pract 2012 Nov 26;14(2):97-130. [doi: 10.1057/dddmp.2012.25]
15. Rogers E. Diffusion of Innovations, 5th Edition. New York: Free Press; 2003.
16. Dennis AR, Kinney ST. Testing media richness theory in the new media: the effects of cues, feedback, and task equivocality. Inform Syst Res 1998 Sep;9(3):256-274. [doi: 10.1287/isre.9.3.256]
17. Facebook. Help centre URL: https://www.facebook.com/unsupportedbrowser [accessed 2018-03-26] [WebCite Cache ID 6yDI6OXx9]
18. Hallett J, Brown G, Maycock B, Langdon P. Changing communities, changing spaces: the challenges of health promotion outreach in cyberspace. Promot Educ 2007;14(3):150-154. [Medline: 18154224]
19. Cassidy E. Gay Men, Social Media and Self-Presentation: Managing Identities in Gaydar, Facebook and Beyond [PhD Thesis]. Brisbane, Austrailia: Queensland University of Technology; 2013. URL: http://eprints.qut.edu.au/61773/ [accessed 2018-03-26] [WebCite Cache ID 6yDIEkTxk]
20. Rubin JD, McClelland SI. 'Even though it's a small checkbox, it's a big deal': stresses and strains of managing sexual identity(s) on Facebook. Cult Health Sex 2015;17(4):512-526. [doi: 10.1080/13691058.2014.994229] [Medline: 25587894]

21. Veale HJ, Sacks-Davis R, Weaver ER, Pedrana AE, Stoové MA, Hellard ME. The use of social networking platforms for sexual health promotion: identifying key strategies for successful user engagement. BMC Public Health 2015 Feb 6;15(1):1. [doi: 10.1186/s12889-015-1396-z]
22. Kite J, Foley BC, Grunseit AC, Freeman B. Please like me: Facebook and public health communication. PLoS One 2016 Sep;11(9):e0162765 [FREE Full text] [doi: 10.1371/journal.pone.0162765] [Medline: 27632172]
23. Rus HM, Cameron LD. Health communication in social media: message features predicting user engagement on diabetes-related Facebook pages. Ann Behav Med 2016 Oct;50(5):678-689. [doi: 10.1007/s12160-016-9793-9] [Medline: 27059761]
24. Capurro D, Cole K, Echavarría MI, Joe J, Neogi T, Turner AM. The use of social networking sites for public health practice and research: a systematic review. J Med Internet Res 2014;16(3):e79 [FREE Full text] [doi: 10.2196/jmir.2679] [Medline: 24642014]
25. Silenzio VM, Duberstein PR, Tang W, Lu N, Tu X, Homan CM. Connecting the invisible dots: reaching lesbian, gay, and bisexual adolescents and young adults at risk for suicide through online social networks. Soc Sci Med 2009 Aug;69(3):469-474 [FREE Full text] [doi: 10.1016/j.socscimed.2009.05.029] [Medline: 19540641]
26. Caldarelli G. Scale-Free Networks: Complex Webs in Nature and Technology. Oxford: Oxford University Press; 2007.
27. Bakshy E, Rosenn I, Marlow C, Adamic L. The role of social networks in information diffusion. In: Proc 21st Int Conf World Wide Web. 2012 Presented at: 21st International Conference on World Wide Web; Apr 16-20, 2012; Lyon, France p. 519-528.
28. Nematzadeh A, Ferrara E, Flammini A, Ahn Y. Optimal network modularity for information diffusion. Phys Rev Lett 2014 Aug 22;113(8):088701. [doi: 10.1103/PhysRevLett.113.088701] [Medline: 25192129]
29. Rieder B. Studying Facebook via data extraction: the Netvizz application. 2013 Presented at: 5th Annual ACM Web Science Conference; May 2-4, 2013; Paris, France.
30. Grant RM, Lama JR, Anderson PL, McMahan V, Liu AY, Vargas L, et al. Preexposure chemoprophylaxis for HIV prevention in men who have sex with men. N Engl J Med 2010 Dec 30;363(27):2587-2599 [FREE Full text] [doi: 10.1056/NEJMoa1011205] [Medline: 21091279]
31. Jacomy M, Venturini T, Heymann S, Bastian M. ForceAtlas2, a continuous graph layout algorithm for handy network visualization designed for the Gephi software. PLoS One 2014 Jun;9(6):e98679 [FREE Full text] [doi: 10.1371/journal.pone.0098679] [Medline: 24914678]
32. Blondel VD, Guillaume J, Lambiotte R, Lefebvre E. Fast unfolding of communities in large networks. J Stat Mech 2008 Oct 09;2008(10):P10008. [doi: 10.1088/1742-5468/2008/10/P10008]
33. Bellazzi R, Diomidous M, Sarkar IN, Takabayashi K, Ziegler A, McCray AT. Data analysis and data mining: current issues in biomedical informatics. Methods Inf Med 2011;50(6):536-544 [FREE Full text] [doi: 10.3414/ME11-06-0002] [Medline: 22146916]
34. Liu B. Sentiment Analysis and Opinion Mining. San Rafael, CA: Morgan & Claypool Publishers; 2012.
35. Nadkarni PM, Ohno-Machado L, Chapman WW. Natural language processing: an introduction. J Am Med Inform Assoc 2011 Sep;18(5):544-551 [FREE Full text] [doi: 10.1136/amiajnl-2011-000464] [Medline: 21846786]
36. Daft RL, Lengel RH. Organizational information requirements, media richness and structural design. Manage Sci 1986 May;32(5):554-571. [doi: 10.1287/mnsc.32.5.554]
37. Brown S. Digital Commons at Cal Poly. 2016 Mar 1. Marketing to Millennials: improving relationships with Millennial consumers through online advertising and social media networking URL: http://digitalcommons.calpoly.edu/marktsp/1
38. Fromm J, Butler C, Dickey C. How to engage Millennials: re-imagining the consumer as a partner, not a target audience, to increase engagement. J Brand Strategy 2015;4(1):36.
39. Dodds PS, Clark EM, Desu S, Frank MR, Reagan AJ, Williams JR, et al. Human language reveals a universal positivity bias. Proc Natl Acad Sci U S A 2015 Feb 24;112(8):2389-2394 [FREE Full text] [doi: 10.1073/pnas.1411678112] [Medline: 25675475]
40. GETPrEPED. URL: https://www.getpreped.ca/ [accessed 2018-03-30] [WebCite Cache ID 6yIythrSn]
41. Resist Stigma. URL: http://resiststigma.com/ [accessed 2018-03-30] [WebCite Cache ID 6yQl8Iakd]
42. Stockman JK, Schwarcz SK, Butler LM, de Jong B, Chen SY, Delgado V, et al. HIV prevention fatigue among high-risk populations in San Francisco. J Acquir Immune Defic Syndr 2004 Apr 01;35(4):432-434. [Medline: 15097163]

Abbreviations

API: application programming interface
CBO: community-based organization
GBM: gay, bisexual, and other men who have sex with men
IRR: incidence rate ratio
PrEP: pre-exposure prophylaxis

Edited by G Eysenbach; submitted 07.06.17; peer-reviewed by J Willoughby, J Bauermeister, M Zlotorzynska, J Kite; comments to author 22.11.17; revised version received 24.01.18; accepted 12.02.18; published 06.04.18.

Please cite as:
Card KG, Lachowsky N, Hawkins BW, Jollimore J, Baharuddin F, Hogg RS
Predictors of Facebook User Engagement With Health-Related Content for Gay, Bisexual, and Other Men Who Have Sex With Men: Content Analysis
JMIR Public Health Surveill 2018;4(2):e38
URL: http://publichealth.jmir.org/2018/2/e38/
doi:10.2196/publichealth.8145
PMID:29625953

Original Paper

Predicting Sexual Behaviors Among Homeless Young Adults: Ecological Momentary Assessment Study

Diane Santa Maria[1], DrPH, RN, MSN; Nikhil Padhye[1], PhD; Yijiong Yang[2], MA; Kathryn Gallardo[2], MSW; Michael Businelle[3], PhD

[1]School of Nursing, University of Texas Health Science Center at Houston, Houston, TX, United States
[2]School of Public Health, University of Texas Health Science Center at Houston, Houston, TX, United States
[3]Department of Family and Preventive Medicine, The University of Oklahoma Health Sciences Center, Oklahoma City, OK, United States

Corresponding Author:
Diane Santa Maria, DrPH, RN, MSN
School of Nursing
University of Texas Health Science Center at Houston
6901 Bertner Ave
Houston, TX, 77030
United States
Phone: 1 713 500 2187
Fax: 1 713 500 2142
Email: diane.m.santamaria@uth.tmc.edu

Related Article:

This is a corrected version. See correction statement: http://publichealth.jmir.org/2018/2/e10806/

Abstract

Background: Homeless youth continue to be disproportionately affected by HIV compared with their housed peers, with prevalence rates as high as 13%. Yet, HIV prevention in this high-risk population has been only marginally effective.

Objective: The aim of this study was to use ecological momentary assessments to examine real-time factors to determine the predictors of sexual activity among homeless youth.

Methods: Youth experiencing homelessness aged between 18 and 24 years were recruited from a drop-in center in Houston, Texas, between August 2015 and May 2016. All the participants received a study-issued mobile phone that prompted brief ecological momentary assessments (EMAs) 5 times a day for 21 days. EMA items assessed near real-time sexual behaviors, cognitions, stress, affect, environmental factors, and environmental circumstances.

Results: Participants (N=66) were predominantly male (41/66, 64%) and black (43/66, 66%) with a median age of 20 years. The mean number of EMAs completed by each participant was 45 out of 105 possible observations. During the study, 70% (46/66) of participants were sexually active and reported condomless sex in 102 of the 137 cases of sexual intercourse (74.5%). In total, 82% (38/46) of the youth who reported having sex during the 3 weeks of data collection also reported engaging in high-risk sexual activities, including having condomless sex (24/46, 53%), having multiple sexual partners on the same day (12/46, 26%), trading sex (7/46, 16%), and sharing needles while injecting drugs (1/46, 3%). Of those, 71% (27/38) were engaged in multiple sexual risk behaviors. The predictive model was based on observations from 66 subjects who reported 137 cases of sexual intercourse over 811 days; sexual orientation, race, mental health, drug use, and sexual urge were included as predictors in the parsimonious generalized linear mixed model selected on the basis of the Akaike information criterion. The estimated odds ratios (ORs) were notable for same-day drug use (OR 8.80, 95% CI 4.48-17.31; *P*<.001) and sexual urge (OR 4.23, 95% CI 1.60-11.28; *P*=.004). The performance of the risk estimator was satisfactory, as indicated by the value of 0.834 for the area under the receiver operating characteristic curve.

Conclusions: Real-time EMA data can be used to predict sexual intercourse among a sample of high-risk, predominately unsheltered homeless youth. Sexual urge and drug use accounts for increased odds of engaging in sexual activity on any given day. Interventions targeting sexual urge and drug use may help predict sexual activity among a population at high risk of HIV.

(JMIR Public Health Surveill 2018;4(2):e39) doi:10.2196/publichealth.9020

KEYWORDS

homeless youth; sexual behaviors; ecological momentary assessment

Introduction

Background

On any given night in the United States, 1.7 to 2.5 million youth under the age of 25 years are homeless [1-3]. Homeless youth are 6 to 12 times more likely to become infected with HIV than housed youth [4], with prevalence rates as high as 13% [5]. Homeless youth are also at greater risk for sexually transmitted infections (STIs) than their housed peers, with 23% of homeless youth reporting having an STI [6]. These high prevalence rates can be partially attributed to the high prevalence of sexual risk behaviors. Homeless youth, become sexually active at an earlier age, are more likely to have multiple sex partners; to trade sex for food, shelter, money, or substances [7-9]; and to use substances before sex. However, they are less likely to use a condom than stably housed youth [10,11]. Youth are also more likely to engage in these high-risk sexual behaviors when they use substances [12]; substance use is high in the homeless youth population [13].

Challenges of Sexual Risk Behavior Research Among Youth Experiencing Homelessness

For decades, prevention research has been challenged by the unsubstantiated belief that homeless youth are beyond help [14]. Prevention interventions specifically targeting homeless youth, though rare, have achieved marginal and temporary improvements in sexual health outcomes (eg, decreased frequency of condomless sex in females) and have been limited to substance-using samples [4,15-18]. However, even modest intervention effects in a marginalized understudied, high-risk population moves the science forward [19]. The disappointing speed at which advancements in prevention research for homeless youth have evolved may be due to the challenges of conducting research among homeless youth [16]. Likewise, our lack of understanding of how real-time factors such as sexual urge, substance use, and stress influence sexual risk behaviors may further stall scientific advancements in this area [20]. The unique experience of homelessness creates significant challenges that need to be addressed such as low concern for STIs, high-risk sexual behaviors [21,22], and the high levels of stress associated with meeting basic needs for food and shelter [11]. Although gender identity, age, race and ethnicity, sexual orientation, educational attainment, and adverse childhood experiences are all linked to sexual risks among homeless youth [23,24], the prevalence of risk behaviors is also elevated by the circumstances experienced before and subsequent to becoming homeless such as sexual abuse and victimization [25-27]. To this end, understanding how the unique extenuating circumstances that precede and extend into homelessness affect one’s thoughts, feelings, and environment, and influence sexual behavior decision making in real time is needed.

Using Ecological Momentary Assessments in Prevention Research Among Youth Experiencing Homelessness

Recall data have a higher potential for bias, neglect intraindividual variability, and do not capture risk and predictive factors as they occur in real-world settings. Ecological momentary assessment (EMA) allows for the examination of within-person variance in risk exposures (ie, where, when, and with whom sexual risk is likely to occur throughout a day) by capturing repeated measures to assess changes in behaviors, cognitions, environmental factors, and symptoms [28,29]. EMA has been used to assess sexual behaviors [30] and drinking in young people [31]. To date, no studies have used EMA to assess whether real-time factors can be used to predict sexual behaviors among homeless youth. EMA may be an effective strategy to gain a better understanding of how real-time thoughts, feelings, and environmental factors affect sexual risk behaviors. To further the science of HIV prevention in homeless youth, strategies must consider the transient nature of being homeless and the varying daily circumstances that influence real-time sexual urge, substance use, stress, and risky decision making. Risk behaviors that are associated with HIV are related to the daily experiences of vulnerability and stress associated with homelessness [27]. The EMA approach is currently the gold standard and most accurate way to measure real-time factors in natural settings [28,32] with high compliance rates (78%) found among youth across 42 studies [33]. High EMA completion rates have been found in other studies on substance using Latino youth (80%) [34], youth in recovery (87%) [35], and youth smokers (88%) [36]. EMA data that are collected at or near the moment when behaviors occur can reduce memory bias and other biases that are associated with retrospective recall measures.

Theoretical Framework

HIV prevention interventions for homeless youth may differ significantly from those for other youth owing to the extenuating circumstances of homelessness. This study was guided by the Risk Amplification Model, which posits that sexual risk behaviors of homeless youth are elevated by the circumstances experienced before and subsequent to homelessness [25-27]. Therefore, there’s a need to understand the role of stress, urge, substance use, and current homeless issues’ impact on HIV risk [37,38].

Purpose

The purpose of this study was to use EMA data to examine real-time factors such as stress, urge, and substance use to determine the predictors of sexual activity among homeless youth. Specifically, the objectives of the study were to determine whether EMA data from among homeless youth can be used to predict sexual risk behaviors.

The primary research question that guided this study was: what are the predictors of sexual intercourse among homeless youth? We hypothesize that real-time factors will predict sexual intercourse. We report the findings of this study using the adapted Strengthening the Reporting of Observational Studies in Epidemiology Checklist for Reporting Ecological Momentary Assessment Studies, which includes the reporting of sampling strategies, measures, schedule, technology used, administration, participant prompting strategy, response rate, and compliance rate [39].

Methods

Participant Recruitment

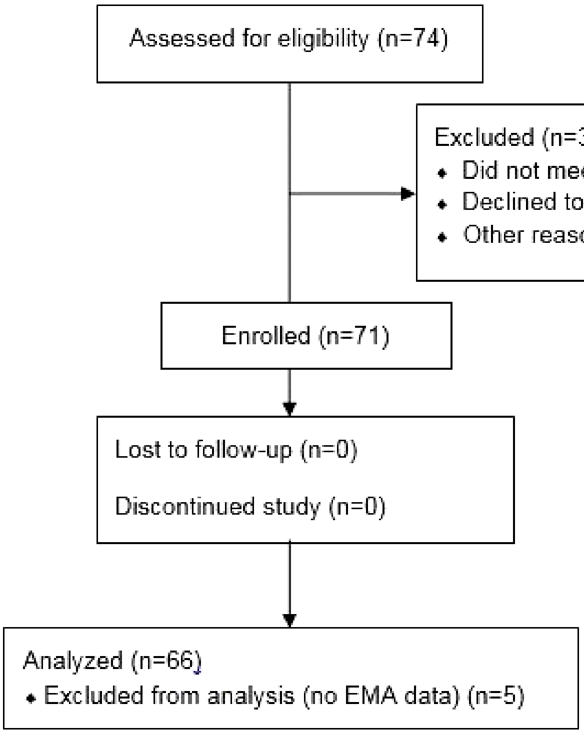
Homeless youth aged between 18 and 24 years were all recruited through study information sessions at shelters and the largest homeless youth drop-in center in Houston, Texas. Flyers were posted at shelters and drop-in centers to advertise for the information sessions. Participants who responded to study advertisements and approached the study team at the drop-in center were provided with the details of the study, and their interest in participating was assessed (see Figure 1). Interested individuals were briefly screened for study eligibility. Participants were included if they were homeless, 18 to 24 years old, English speaking, and able to participate for the duration of the study period (ie, not planning to move out of the county during the study). Homelessness was defined as sleeping on the streets, in a place not meant for human habitation, in a shelter, in a hotel or motel, or with someone with whom they could not stay for more than 30 days (ie, couch surfing). Individuals were excluded if they had very low literacy, owing to the need to be able to read and understand EMAs unassisted throughout the study [40]. Low literacy was defined as a score of less than 4 on the Rapid Estimate of Adult Literacy in Medicine-Short Form [40,41], an interviewer-administered checklist in which individuals are asked to read and pronounce 9 common medical terms. Individuals who pronounce ≥4 words correctly are considered to be reading at >6th grade reading level.

Study Procedures

Details of the study were discussed with eligible youth, who then provided written informed consent, witnessed by the study staff, and received a study summary and a copy of the informed consent document. After the initial eligibility screening, participants completed an audio-assisted baseline survey on an Apple iPad using Qualtrics. The baseline survey took approximately 30 min to complete and assessed demographics such as gender identity, age, race/ethnicity, sexual orientation, educational attainment, and adverse childhood experiences. Participants were then provided with a study-issued mobile phone, with instructions for using the phone and accessing and completing the EMAs over the 21-day study duration. Each participant was asked to indicate his/her normal waking hours to ensure that EMA prompts would not wake the participant. Snacks and beverages were provided and breaks were encouraged as needed. Participants received a US $20 gift card for completing the baseline survey and were provided with bus tickets or METRO pass when needed to cover the cost of local transportation on baseline and exit study visits.

Study staff contacted participants on their study phone to schedule the final study visit 21 days after the initial study visit. Participants were asked to meet at a local drop-in center, a shelter, or a local library to complete an exit survey, return the study mobile phone, and receive grocery store gift cards. Upon returning the study smartphone at the final visit, participants received up to $95 in gift cards. The amount of compensation depended upon the percentage of random and daily EMAs completed. Specifically, participants who completed 49.5% (52/105) to 75.2% (79/105) of EMAs received a $50 gift card, those who completed 76.2% (80/105) to 88.6% (93/105) of EMAs received a $75 gift card, and those who completed 89.5% (94/105) or more of EMAs received a $95 gift card. Those who completed <50% of assessments received a $20 gift card for returning the phone. This incentive structure was explained to all participants during the informed consent process. Participants were able to access their current compensation level throughout the study period through the study-issued phone interface.

Figure 1. CONSORT (Consolidated Standards of Reporting Trials) E-HEALTH Flow Diagram. EMA: ecological momentary assessment.



Baseline Measures

We assessed age, race/ethnicity, sexual orientation, and mental illness at baseline. To assess race/ethnicity, youth were asked if they identified as black, white, Asian, Hispanic, American Indian, multiracial, or something else. We created a category called "Other" that included those who identified as American Indian or Alaska Native, multiracial, or something else. Sexual orientation was measured by asking youth if they identified as heterosexual, gay, lesbian, bisexual, or something else. Mental illness was assessed by asking youth if they had ever been diagnosed with attention deficit-hyperactivity disorder (ADD/ADHD), depression, bipolar disorder, psychosis, schizophrenia, oppositional defiant disorder, conduct disorder, or posttraumatic stress disorder (PTSD).

Ecological Momentary Assessment Measures Schedule

EMAs were used to collect data in near real time using study-provided mobile phones. The EMA methodology used is similar to that developed by Shiffman, Stone, and colleagues [29,42,43] and has been used by our study team in multiple studies [44-46]. EMAs were prompted 5 times a day. Time-based sampling (daily diary) and random sampling EMAs were completed. The phone audibly and visually cued each assessment for 30 seconds. If the participant did not respond after 3 prompts, the assessment was recorded as missed. Daily diary assessments were prompted once every day 30 min after the participant's indicated normal waking time. Questions referred to the previous 24 hours and queried about risk behaviors. The daily EMAs assessed items such as sheltering, and engaging in sexual activity, substance use, and alcohol consumption. Shelter day referred to youth who had spent the night in a shelter versus on the street or staying with someone temporarily. Questions such as "Did you have sex yesterday?," "Did you trade sex yesterday?," "Did you drink any alcoholic beverages yesterday?," and "Did you view pornography yesterday?" were used to assess behaviors. If the participant answered "yes" to the sexual activity item, the participant was prompted to indicate the type of sexual activity (eg, oral, vaginal, or anal), whether a condom was used, and the number, and gender of the sexual partner or partners. The 4-item Perceived Stress Scale was used to measure stress. Scores were summed, with higher scores indicating more stress. Daily diary assessments took less than 5 min to complete. For the main outcome of sexual activity, each incident of sexual activity as indicated on the daily EMA, was considered a positive case.

Random assessments were scheduled to occur randomly in 4 epochs during each participant's normal waking hours. Random EMAs took approximately 2 min to complete. In the random EMAs, participants rated their current affect by indicating the extent to which they felt irritable, happy, content, frustrated or angry, sad, worried, miserable, restless, stressed, hostile, and calm. Behaviors were assessed by asking about substance use and alcohol use. Sexual urge was measured by asking youth if they were feeling a strong urge to have sex, use drugs or alcohol, or steal.

Technology and Hardware

The Samsung Galaxy Light mobile phone with the Android 4.2 operating system was used to send EMAs. Participants could call and receive calls from research staff through the smartphone free of charge.

Data Analysis

R, package lme4, and package pROC (R project) were used in the analyses [47-49]. Generalized linear mixed models (GLMMs) were used to model the high-frequency longitudinal EMA data with a logistic link used for adaptation to the binary outcome of sexual intercourse. A random intercept was added to account for correlated observations within subject. The identified predictors of the dependent variables included EMA data that preceded the case behavior and assessed real-time sexual urge, substance use, and stress up to 24 hours before the occurrence of sexual intercourse. To maximize the use of available data, we built models in 3 steps. Missing data were not estimated but were ignored to avoid the introduction of bias due to the estimation method. Time-invariant predictors (ie, demographic data) were included in the first step. The Akaike information criterion (AIC) [50] was used to systematically eliminate variables that were not predictive with backward selection. Daily event data were introduced in the second step, and predictors were similarly eliminated using AIC for model selection. Random EMA data were summarized by day to match the collection frequency of the outcome before being introduced in the third step of modeling. Model performance was assessed with a receiver operating characteristic (ROC) curve and with cross-validated sensitivity and specificity. Cross-validation was accomplished over the course of 100 runs with 80/20 random splits of training and test sets [51]. For measuring sensitivity and specificity, the model estimates of the probabilities of sexual intercourse events were converted to binary predictions of sex events (yes or no) with the choice of a decision threshold, (eg, sex event predicted when the probability of sexual intercourse exceeded 0.3). The decision threshold was chosen to provide a reasonable balance between true positive and false positive rate, which was measured with the likelihood ratio.

Results

Sample Characteristics

The mean age of the sample was 21.2 years, with 55% (36/66) aged between 21 and 24 years. The majority of participants were male (41/66, 62%), black (43/66, 65%) or other race (16/66, 24%), and a minority (14/66, 21%) identified as lesbian, gay, bisexual, transgender, or questioning (Table 1). The median age at the onset of homelessness was 16.9 years. Only 13.4% (108/806) of days were shelter days, with most days being on the streets (293/806 days) or staying with someone temporarily (405/806 days). At baseline, 20% (13/66) of participants indicated a history of having an STI diagnosis. Among the total sample, 73% (48/66) reported at least 1 mental health diagnosis (ADHD, bipolar disorder, depression, oppositional defiant disorder, conduct disorder, psychosis, schizophrenia, or PTSD). Of these, 11% (7/66) had a diagnosis of psychosis, 21% (14/66) reported PTSD diagnosis, and 46% (30/66) reported a diagnosis of bipolar disorder, manic depression, or depression. More than half of the participants reported having a diagnosis of ADD/ADHD.

Response and Compliance Findings

We received EMA data from 66 of 71 recruited participants, indicating a 93% participation rate (Table 2). Table 2 provides the number and percentage of surveys that endorsed the variable, number of EMAs included in each count, and the number of participants who provided EMA data by variable. The mean number of EMAs provided by each participant was 45 observations—that is, 13.0 out of 21 (61.9%) of daily EMAs and 33.6 out of 84 (40%) of random EMAs.

The average daily EMA compliance rate of 61.9% conservatively assumes that all participants received all of the EMA over the 21 days for a total of 105 (Table 3). However, due to sheltering instability among the sample, participants reported frequent loss of phone battery charge due to having no available electrical outputs. This likely decreased the actual number of EMAs received. For the random data, each participant had the potential to receive 4 random surveys per day for 21 days. Conservatively, the average number of random EMAs per participant was 33.6 out of 84, indicating a random EMA average compliance rate of 40.0%. The compliance rate of daily EMAs across all was higher than the compliance rate of random EMAs. Participants who identified as other race, female, and older than 20 years of age completed more random EMAs than their counterparts. Random EMA completion rates were similar across sexual orientations.

Table 1. Participant demographics and sexual behaviors. LGBT: lesbian, gay, bisexual, and transgender; ADD/ADHD: attention deficit-hyperactivity disorder; ODD: oppositional defiance disorder; CD: conduct disorder; PTSD: posttraumatic stress disorder.

Demographics	Participants	Condomless sex[a]	Multiple sexual partners[a]	Trade sex[a]	Substance use[a]	Alcohol use[a]
Participants, n	66	35	17	11	40	21
Age, years, n (%)						
≤20	30 (46)	22 (63)	8 (47)	7 (64)	20 (50)	7 (33)
>20	36 (55)	13 (37)	9 (53)	4 (36)	20 (50)	14 (67)
Gender identity, n (%)						
Male	41 (62)	19 (54)	11 (65)	4 (36)	25 (63)	11 (52)
Female	24 (36)	15 (43)	6 (35)	7 (64)	14 (35)	9 (43)
Other[b]	1 (2)	1 (3)	0 (0)	0 (0)	1 (3)	1 (5)
Race, n (%)						
Black	43 (65)	23 (66)	13 (77)	8 (73)	25 (63)	14 (67)
White	7 (11)	2 (6)	1 (6)	0 (0)	4 (10)	2 (10)
Other[c]	16 (24)	10 (29)	3 (18)	3 (27)	11 (28)	5 (24)
Ethnicity, n (%)						
Hispanic	8 (12)	4 (11)	2 (12)	2 (18)	4 (10)	3 (14)
Sexual orientation, n (%)						
Heterosexual	52 (79)	24 (69)	10 (59)	3 (27)	28 (70)	11 (52)
LGBT	14 (21)	11 (31)	7 (41)	8 (73)	12 (30)	10 (48)
Foster History, n (%)	22 (34)	39 (38)	14 (45)	10 (59)	85 (41)	24 (55)
Mental health diagnosis, n (%)						
ADD or ADHD	37 (56)	23 (66)	12 (71)	7 (64)	26 (65)	14 (67)
Bipolar or manic	30 (46)	18 (51)	9 (53)	7 (64)	19 (48)	11 (52)
Depression	30 (46)	18 (51)	7 (20)	6 (55)	20 (50)	12 (57)
ODD or CD	13 (20)	9 (26)	7 (20)	3 (27)	10 (25)	6 (29)
Psychosis	7 (11)	3 (9)	2 (6)	2 (18)	4 (10)	2 (10)
PTSD	14 (21)	9 (26)	5 (14)	3 (27)	8 (20)	4 (19)

[a] Number of participants who engaged in the specified behavior at least once.

[b] Other: transgender, nonbinary gender.

[c] Other: American Indian or Alaska Native; Multiracial; Something else.

Table 2. Ecological momentary assessment (EMA) and participant count and frequency by variable.

Variable	EMA surveys endorsing[a], n (%)	Participants, n (%)[b]
Sexual activity	137 (16.9)	46 (70)
Multiple sexual partner	31 (22.1)	17 (26)
Condomless sex	102 (74.5)	35 (53)
Trade sex	17 (12.4)	11 (17)
Substance use	210 (26.1)	40 (61)
Sexual urge	124 (15.5)	44 (67)
Alcohol use	44 (6.7)	21 (32)
High stress day	235 (29.0)	47 (71)
Pornography use	152 (17.7)	43 (65)
Shelter day	108 (13.4)	4 (6)

[a] Number and percentage of the total number of surveys that endorsed this behavior.

[b] Number of participants of the total 66 participants who engaged in the behavior at least once.

Table 3. Ecological momentary assessment (EMA) compliance rates by demographics. LGBT: lesbian, gay, bisexual, and transgender.

Variable	Total EMA (n=105), mean (SD)	Daily EMA (n=21), mean (SD)	Random EMA (n=84), mean (SD)
Whole sample	46.7 (28.52)	13.03 (6.32)	33.6 (22.8)
Race			
Black	45.47 (28.72)	12.91 (6.51)	32.56 (22.8)
White	46.71 (29.91)	12.86 (7.08)	33.86 (23.46)
Other	50.6 (28.95)	13.44 (5.89)	23.88 (23.88)
Gender			
Male	42.08 (24.83)	12.27 (6.16)	29.55 (19.45)
Female	52.88 (32.60)	14 (6.51)	38.88 (26.32)
Other	89 (-)	21 (-)	68 (-)
Age, years			
≤20	43.83 (26.81)	12.43 (5.85)	31.03 (21.61)
>20	49.17 (30.00)	13.53 (6.74)	35.64 (23.77)
Sexual orientation			
Heterosexual	46.82 (28.36)	12.96 (6.25)	33.65 (22.66)
LGBT	46.64 (30.17)	13.29 (6.84)	33.36 (24.05)

Table 4. Generalized linear mixed models (GLMM) coefficients and odds ratios for predictors of sexual intercourse. OR: odds ratio; LGBT: lesbian, gay, bisexual, and transgender; PTSD: posttraumatic stress disorder.

Variable	Coefficient B	SE	OR	Z	*P* value	95% CI of OR
Fixed effects						
Intercept	−2.846	0.576	0.06	−4.944	<.001	0.019-0.180
Sexual orientation (LGBT)[a]	0.8703	0.4061	2.388	2.143	.03	1.077-5.290
Race (white)[a]	−0.7501	0.6724	0.472	−1.116	.27	0.127-1.763
Race (other)[a]	0.9205	0.4207	2.511	2.188	.03	1.101-5.733
Psychosis[a]	1.4716	0.6195	4.356	2.376	.02	1.293-14.690
PTSD[a]	−1.6613	0.4681	0.190	−3.549	<.001	0.076-0.475
Drug use	2.1748	0.3445	8.800	6.313	<.001	4.476-17.309
Sexual urge	1.4431	0.4999	4.234	2.887	.004	1.589-11.280

[a] Reference group is black, heterosexual youth without mental illness.

Ecological Momentary Assessment

In total, 66 participants completed 860 daily EMAs (Table 1). We analyzed the daily EMAs descriptively and found that work and school days among the sample were low (6% each). Of the 66 participants, 38 (58%) reported engaging in high-risk sexual behaviors during the data collection period, including having condomless sex, having multiple sexual partners in the same day, trading sex, or sharing needles to inject drugs. Of those 38 youth, 26 (71%, 26/38) had engaged in more than one of the risk behaviors during the study. Substance use rates were also high, 62% (40/66) and 32% (21/66) reported using drugs and alcohol, respectively.

Although 70% (51/66) of the participants were sexually active during the study, 53% (35/66) reported condomless sex accounting for 102 incidences or 75% of sexual intercourse incidences. Additionally, 26% (17/66) had sex with more than one person in a day, 16% (11/66) engaged in trade sex, and 18% (12/66) used pornography during the 3-week EMA period. Pornography use appeared to be a cluster behavior with other risk behaviors occurring on the same day as 40% of condomless sex days, 68% of multiple sexual partner days, 65% of trade sex days, and 46% of substance use days. Sexual urge was experienced by 67% of the participants and was reported on 124 or 15% of daily EMA study days. Higher rates of sexual intercourse were reported on high sexual urge days and drug use days.

Risk Estimator for Sexual Activity Days

The final predictive model for sexual intercourse included both between- and within-subject variables: race, sexual orientation, mental illness, drug use, and sexual urge in the parsimonious GLMM selected on the basis of the AIC (Table 4). The risk estimator was based on observations from 66 participants over 860 days and included 137 days of sexual intercourse. Of note, not all predictors had significant *P* values, factors needed to be included to minimize the AIC, indicating that they were not ignorable in predicting substance use. The estimated odds ratios (ORs) for the within-subject predictors were notable for 2 states; drug use (OR 8.80, *P*<.001) and sexual urge (OR 4.23, *P*=.004). The odds of having sex increase 8.8 times on days when youth use drugs after adjusting for sexual urge and other predictors. The odds of having sex increase 4.2 times per unit increase in sexual urge, after adjusting for drug use and other predictors.

The performance of the risk estimator was very good, as indicated by the value of 0.834 for the area under the ROC curve [52]. The cross-validation run over 100 times with randomized 80/20 splits of the data into training and test sets resulted in a mean sensitivity of 0.17 and specificity of 0.96 with a likelihood ratio of 4.45 for a decision threshold of *P*=.48 (ie, sex predicted if predicted probability exceeds 0.5). For a decision threshold of *P*=.20 (ie, sex predicted if predicted probability exceeds 0.20), the sensitivity and specificity were 0.637 and 0.832, respectively with a likelihood ratio of 3.80.

Finally, we compared the model using only traits as predictors versus the addition of drug use and sexual urge. The area under the ROC curve dropped from 0.834 to 0.683, indicating that the model improves with the state variables as values below 0.7 are often considered unsatisfactory. Additionally, using a decision threshold of *P*=.20 to predict sex events, the cross-validated sensitivity dropped to 50.1%, the specificity dropped to 73.3%, and the likelihood ratio halved from 3.8 to 1.9.

Discussion

The findings presented here represent a predominantly unsheltered and unstably housed sample of young adults with high-risk sexual behaviors and high rates of substance use. Sexual intercourse was predicted by both between- and within-subject variables, including real-time drug use and sexual urge. Subgroups of homeless youth emerged as higher risk for sexual intercourse. The odds of sexual intercourse were highest among nonwhite, other race youth, those that identify as lesbian, gay, bisexual, and transgender (LGBT), and those who have ever been diagnosed with PTSD. Therefore, special attention is needed to address HIV risk reduction within minority youth and those with mental illness.

Our findings suggest that the majority of the homeless youth in this study were sexually active and they primarily engage in condomless sex, placing them at heightened risk for HIV, STIs, and pregnancy. While previous studies have demonstrated the high rates of sexual risk behaviors among homeless youth [8,11,53], the majority of those studies used a cross-sectional study design and assessed the occurrence of sexual risk behaviors either by recall of behaviors within the past 1 to 3 months or by assessing condom use at last sex. These studies reported rates of condomless sex that ranged from 40% to 70% [54]. By using EMA in a high frequency longitudinal study design, we assessed behaviors in near real-time by asking participants to record sexual risk behaviors that occurred within the past 24 hours. This methodology has been shown to reduce recall biases that are commonly associated with cross-sectional retrospective recall measures. By asking youth to report condomless sex per instance of sexual intercourse in near real-time, we found that rates of condom use were much lower than those reported in studies using recall measures or assessing condom use during the last sexual intercourse.

Although characteristics of homeless youth suggest higher risk subgroups, one’s real-time urges and drug use also influence the odds of engaging in sexual intercourse. Youth experiencing homelessness are at higher risk of engaging in sexual intercourse on the days they use substances. In alignment with previous literature [11,13], we found that homeless youth engage in high rates of substance use. Approximately 60% of participants reported using substances and nearly one-third of participants reported using alcohol during the study period. More importantly, we found that homeless youth were more likely to engage in sexual intercourse on the days they reported using substances. Previous studies demonstrated that the use of drugs and alcohol can impair homeless youth’s sexual health decision-making capabilities, thus increasing their risk for engaging in sexual risk behaviors and acquiring HIV [12,21,55]. Additionally, the Risk Amplification Model suggests that homeless youth form high-risk social networks. Therefore, when youth reported unstable housing days, these relationships may represent high-risk social networks that may contribute to the risk for substance use.

We found that youth are at higher odds of engaging in sexual intercourse on days when they experience sexual urges. Kennedy and colleagues found that experiencing feelings of intense sexual arousal influenced homeless youth’s decision to engage in unprotected sex [22]. Thus, our results align with the research that experiencing sexual urges is a risk factor for subsequently engaging in sexual intercourse and that sexual urge can be detected in real time using EMA.

The results of our study also indicated that pornography use was common among homeless youth, with approximately 65% of participants reported viewing pornography at least once during the study period. Interestingly, viewing pornography appeared to be clustered with other risk behaviors, including having condomless sex, trading sex, and having multiple sexual partners on the same day. Our findings support previous research examining the impact on sexual risk behaviors of pornography use among youth and young adults [56,57]. For example, Braun-Courville and Rojas found that youth who viewed pornography were more likely to engage in sexual risk behaviors [57]. This represents an understudied area among homeless youth that warrants more attention.

Significance of the Findings

This is the first study to use EMA data to predict the likelihood of engaging in sexual intercourse in near real time among a high-risk, hard-to-reach, homeless youth population. Using EMA, we found that 26% (17/66) of youth experiencing homelessness had multiple sexual partners in a day, and about 53% (35/66) engaged in condomless sex which likely contributes to the high rate of HIV and STIs among homeless youth. Additionally, the EMA and statistical analysis methods used here are potentially applicable to other hard-to-reach populations and can be used to predict other risk behaviors that occur with frequency and are potentially affected by real-time cognitions and behaviors.

Implications for Research and Prevention Interventions

This study demonstrates that it is possible to predict days when youth are at higher risk for engaging in sexual intercourse. To this end, it may be possible to develop just-in-time interventions that can disrupt the progression from drug use and sexual urge to engaging in sexual risk behaviors by addressing drug use and identifying skills to manage sexual urge in ways that reduce sexual risks. If we can predict days when youth are at higher risk of having sexual intercourse, we can design and test safer sex promoting motivational messaging that can be delivered at the time of heightened risk and have the potential to enhance safer sexual behavior decision making. For example, since these data revealed that 75% of sexual acts were condomless, messages encouraging condom use could be sent to participants on the days when youth report sexual urge or drug use.

Limitations

The findings represented here may not reflect other possible real-time predictors that were not measured. Although we constructed a comprehensive EMA survey based on extensive formative research [58-60] that included variables indicated in the literature to affect sexual behaviors, other variables may also influence real-time sexual risk. That said, using EMA to predict risk behaviors is a relatively novel scientific method that does not have defined guidelines for best practices for measurement, implementation, or analysis, particularly among vulnerable populations such as homeless youth. In so far as the EMA approach is an emerging science, the measures used to assess real-time factors have not yet been psychometrically validated. The temporality of the data is another limitation. EMA improves on the ability to assess subsequent behaviors from real-time measures. However, participants reported their sex behaviors from the preceding day. Drug use and sexual urge EMA variables were calculated for the preceding day to align with the sexual behavior variable. Therefore, we cannot unequivocally conclude, on the basis of these data, that drug use and urge preceded sexual intercourse on a given day. Finally, though we used statistical methods to cross-validate the predictive model, a subsequent study in a new population of homeless youth to test the model would provide further validation.

Conclusions

High frequency longitudinal EMA data that assess real-time factors can be used to predict sexual intercourse. This kind of data and analyses can inform the design of just-in-time adaptive interventions that could be delivered using mobile phones to deliver health promoting and motivational messaging at the time of heightened risk.

Acknowledgments

This study was funded by the UTHealth PARNTERS Research Grant and the University of California San Francisco Center for AIDS Prevention Studies supported through an award from the National Institute of Child Health and Human Development (R25HD045810). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. The authors are especially grateful to the youth who participated in this study.

Authors' Contributions

DSM conceived the study design and manuscript topic, developed the EMA measures, supervised data collection and data analysis, and drafted the manuscript. NP and YY conducted the data analysis and contributed to the manuscript development. MB supervised the EMA development, contributed to the EMA measures, study design, interpretation of the results, and manuscript writing. KG assisted with data collection and manuscript development.

Conflicts of Interest

None declared.

References

1. Bassuk EL, Murphy C, Coupe N, Kenney R, Beach C. American Institutes for Research. 2014. America's youngest outcasts 2014: state report card on child homelessness URL: https://www.air.org/sites/default/files/downloads/report/Americas-Youngest-Outcasts-Child-Homelessness-Nov2014.pdf [accessed 2018-03-30] [WebCite Cache ID 6yIQUdKD7]
2. Bassuk E, Murphy C, Coupe N, Kenney R. The National Center on Family Homelessness. 2011. America's youngest outcasts 2010: state report on child homelessness URL: http://www.hartfordinfo.org/issues/wsd/Homelessness/NCFH_AmericaOutcast2010_web.pdf [accessed 2018-03-30] [WebCite Cache ID 6yIRlAbdo]
3. Bassuk EL. Ending child homelessness in America. Am J Orthopsychiatry 2010 Oct;80(4):496-504. [doi: 10.1111/j.1939-0025.2010.01052.x] [Medline: 20950290]
4. Rotheram-Borus MJ, Song J, Gwadz M, Lee M, Van Rossem R, Koopman C. Reductions in HIV risk among runaway youth. Prev Sci 2003 Sep;4(3):173-187. [Medline: 12940468]
5. Pfeifer RW, Oliver J. A study of HIV seroprevalence in a group of homeless youth in Hollywood, California. J Adolesc Health 1997 May;20(5):339-342. [doi: 10.1016/S1054-139X(97)00038-4] [Medline: 9168380]
6. Carmona J, Slesnick N, Guo X, Letcher A. Reducing high risk behaviors among street living youth: outcomes of an integrated prevention intervention. Child Youth Serv Rev 2014 Aug 01;43:118-123 [FREE Full text] [doi: 10.1016/j.childyouth.2014.05.015] [Medline: 25104870]
7. Ray N. The National Gay and Lesbian Task Force. 2006. Lesbian, gay, bisexual and transgender youth: an epidemic of homelessness URL: http://www.thetaskforce.org/static_html/downloads/reports/reports/HomelessYouth.pdf [accessed 2018-03-30] [WebCite Cache ID 6yIS2EYAx]
8. Beech B, Myers L, Beech D, Kernick NS. Human immunodeficiency syndrome and hepatitis B and C infections among homeless adolescents. Semin Pediatr Infect Dis 2003 Jan;14(1):12-19. [doi: 10.1053/spid.2003.127212] [Medline: 12748917]
9. Gangamma R, Slesnick N, Toviessi P, Serovich J. Comparison of HIV risks among gay, lesbian, bisexual and heterosexual homeless youth. J Youth Adolesc 2008;37(4):456-464 [FREE Full text] [doi: 10.1007/s10964-007-9171-9] [Medline: 18607514]
10. Thompson S, Pillai V. Determinants of runaway episodes among adolescents using crisis shelter services. Int J Soc Welfare 2006;15(2):142-149. [doi: 10.1111/j.1468-2397.2006.00370.x]
11. Edidin J, Ganim Z, Hunter S, Karnik N. The mental and physical health of homeless youth: a literature review. Child Psychiatry Hum Dev 2012 Jun;43(3):354-375. [doi: 10.1007/s10578-011-0270-1] [Medline: 22120422]
12. Tapert SF, Aarons GA, Sedlar GR, Brown SA. Adolescent substance use and sexual risk-taking behavior. J Adolesc Health 2001 Mar;28(3):181-189. [Medline: 11226840]
13. Kemp PA, Neale J, Robertson M. Homelessness among problem drug users: prevalence, risk factors and trigger events. Health Soc Care Community 2006 Jul;14(4):319-328. [doi: 10.1111/j.1365-2524.2006.00624.x] [Medline: 16787483]
14. Slesnick N. Bleak and hopeless no more. Engagement of reluctant substance-abusing runaway youth and their families. J Subst Abuse Treat 2000;19(3):215-222. [Medline: 11027890]
15. Naranbhai V, Abdool KQ, Meyer-Weitz A. Interventions to modify sexual risk behaviours for preventing HIV in homeless youth. Cochrane Database Syst Rev 2011 Jan 19(1):CD007501 [FREE Full text] [doi: 10.1002/14651858.CD007501.pub2] [Medline: 21249691]

16. Altena A, Brilleslijper-Kater S, Wolf J. Effective interventions for homeless youth: a systematic review. Am J Prev Med 2010 Jun;38(6):637-645. [doi: 10.1016/j.amepre.2010.02.017] [Medline: 20494240]
17. Slesnick N, Kang M. The impact of an integrated treatment on HIV risk behavior among homeless youth: a randomized controlled trial. Journal of behavioral medicine 2008;31(1):45-59.
18. Slesnick N, Prestopnik JL. Ecologically based family therapy outcome with substance abusing runaway adolescents. J Adolesc 2005 Apr;28(2):277-298 [FREE Full text] [doi: 10.1016/j.adolescence.2005.02.008] [Medline: 15878048]
19. Slesnick N, Dashora P, Letcher A, Erdem G, Serovich J. A review of services and interventions for runaway and homeless youth: moving forward. Child Youth Serv Rev 2009 Jul;31(7):732-742 [FREE Full text] [doi: 10.1016/j.childyouth.2009.01.006] [Medline: 20161294]
20. Heerde J, Scholes-Balog K, Hemphill S. Associations between youth homelessness, sexual offenses, sexual victimization, and sexual risk behaviors: a systematic literature review. Arch Sex Behav 2015 Jan;44(1):181-212. [doi: 10.1007/s10508-014-0375-2] [Medline: 25411128]
21. Booth RE, Zhang Y, Kwiatkowski CF. The challenge of changing drug and sex risk behaviors of runaway and homeless adolescents. Child Abuse Negl 1999 Dec;23(12):1295-1306. [Medline: 10626612]
22. Kennedy D, Brown R, Morrison P, Vie L, Ryan G, Tucker J. Risk evaluations and condom use decisions of homeless youth: a multi-level qualitative investigation. BMC public health 2015;15(1):62. [Medline: 25636862]
23. Felitti VJ, Anda RF, Nordenberg D, Williamson DF, Spitz AM, Edwards V, et al. Relationship of childhood abuse and household dysfunction to many of the leading causes of death in adults. The Adverse Childhood Experiences (ACE) Study. Am J Prev Med 1998 May;14(4):245-258. [Medline: 9635069]
24. Slesnick N, Zhang J, Brakenhoff B. Homeless youths' caretakers: the mediating role of childhood abuse on street victimization and housing instability. Soc Work 2016 Apr;61(2):147-154. [Medline: 27180525]
25. Thrane LE, Yoder KA, Chen X. The influence of running away on the risk of female sexual assault in the subsequent year. Violence Vict 2011;26(6):816-829. [Medline: 22288098]
26. Tyler KA, Johnson KA. Trading sex: voluntary or coerced? the experiences of homeless youth. J Sex Res 2006 Aug;43(3):208-216. [doi: 10.1080/00224490609552319] [Medline: 17599243]
27. Whitbeck LB, Hoyt DR, Yoder KA. A risk-amplification model of victimization and depressive symptoms among runaway and homeless adolescents. Am J Community Psychol 1999 Apr;27(2):273-296. [doi: 10.1023/A:1022891802943] [Medline: 10425702]
28. Stone A, Shiffman S. Capturing momentary, self-report data: a proposal for reporting guidelines. Ann Behav Med 2002;24(3):236-243. [doi: 10.1207/S15324796ABM2403_09] [Medline: 12173681]
29. Shiffman S. Ecological momentary assessment (EMA) in studies of substance use. Psychol Assess 2009 Dec;21(4):486-497 [FREE Full text] [doi: 10.1037/a0017074] [Medline: 19947783]
30. Shrier LA, Feldman HA, Black SK, Walls C, Kendall AD, Lops C, et al. Momentary affective states surrounding sexual intercourse in depressed adolescents and young adults. Arch Sex Behav 2012 Oct;41(5):1161-1171. [doi: 10.1007/s10508-011-9787-4] [Medline: 21755382]
31. Wright C, Dietze P, Crockett B, Lim M. Participatory development of MIDY (mobile intervention for drinking in young people). BMC Public Health 2016 Feb 24;16:184 [FREE Full text] [doi: 10.1186/s12889-016-2876-5] [Medline: 26911299]
32. Stone A, Shiffman S, Atienza A, Nebeling L. The Science of Real-Time Data Capture: Self-Reports in Health Research. New York, NY: Oxford University Press; 2007.
33. Wen C, Schneider S, Stone S, Spruijt-Metz D. Compliance with mobile ecological momentary assessment protocols in children and adolescents: a systematic review and meta-analysis. J Med Internet Res 2017 Apr 26;19(4):e132 [FREE Full text] [doi: 10.2196/jmir.6641] [Medline: 28446418]
34. Comulada W, Lightfoot M, Swendeman D, Grella C, Wu N. Compliance to cell phone-based EMA among Latino youth in outpatient treatment. J Ethn Subst Abuse 2015;14(3):232-250 [FREE Full text] [doi: 10.1080/15332640.2014.986354] [Medline: 26114764]
35. Dennis M, Scott C, Funk K, Nicholson L. A pilot study to examine the feasibility and potential effectiveness of using smartphones to provide recovery support for adolescents. Subst Abus 2015;36(4):486-492 [FREE Full text] [doi: 10.1080/08897077.2014.970323] [Medline: 25310057]
36. Hoeppner BB, Kahler CW, Gwaltney CJ. Relationship between momentary affect states and self-efficacy in adolescent smokers. Health Psychol 2014 Dec;33(12):1507-1517 [FREE Full text] [doi: 10.1037/hea0000075] [Medline: 25020151]
37. Bartholomew LK, Mullen PD. Five roles for using theory and evidence in the design and testing of behavior change interventions. J Public Health Dent 2011;71 Suppl 1:S20-S33. [Medline: 21656946]
38. Eldredge K, Markham C, Ruiter R, Kok G, Parcel G. Planning Health Promotion Programs: An Intervention Mapping Approach. San Francisco, CA: Jossey-Bass; 2016.
39. Liao Y, Skelton K, Dunton G, Bruening M. A systematic review of methods and procedures used in ecological momentary assessments of diet and physical activity research in youth: an adapted STROBE checklist for reporting EMA studies (CREMAS). J Med Internet Res 2016 Dec 21;18(6):e151 [FREE Full text] [doi: 10.2196/jmir.4954] [Medline: 27328833]

40. Shea J, Beers B, McDonald V, Quistberg A, Ravenell K, Asch D. Assessing health literacy in African American and Caucasian adults: disparities in rapid estimate of adult literacy in medicine (REALM) scores. Fam Med 2004 Sep;36(8):575-581 [FREE Full text] [Medline: 15343419]
41. Murphy P, Davis T, Long S, Jackson R, Decker B. Rapid estimate of adult literacy in medicine (REALM): a quick reading test for patients. J Read 1993;37(2):124-130 [FREE Full text] [doi: 10.2307/40033408]
42. Shiffman S, Hickcox M, Paty JA, Gnys M, Richards T, Kassel JD. Individual differences in the context of smoking lapse episodes. Addict Behav 1997;22(6):797-811. [Medline: 9426798]
43. Stone AA, Schwartz JE, Neale JM, Shiffman S, Marco CA, Hickcox M, et al. A comparison of coping assessed by ecological momentary assessment and retrospective recall. J Pers Soc Psychol 1998 Jun;74(6):1670-1680. [Medline: 9654765]
44. Watkins K, Reagan S, Nguyen N, Businelle M, Kendzor D, Lam C, et al. Advancing cessation research by integrating EMA and geospatial methodologies: associations between tobacco retail outlets and real-time smoking urges during a quit attempt. Nicotine Tob Res 2013;16(Suppl 2):S93-101. [Medline: 24057995]
45. Businelle M, Ma P, Kendzor D, Reitzel L, Chen M, Lam C, et al. Predicting quit attempts among homeless smokers seeking cessation treatment: an ecological momentary assessment study. Nicotine Tob Res 2014;16(10):1371-1378. [Medline: 24893602]
46. Reitzel LR, Kendzor DE, Nguyen N, Regan SD, Okuyemi KS, Castro Y, et al. Shelter proximity and affect among homeless smokers making a quit attempt. Am J Health Behav 2014 Mar;38(2):161-169 [FREE Full text] [doi: 10.5993/AJHB.38.2.1] [Medline: 24629545]
47. R Development Core Team. Vienna, Austria: Foundation for Statistical Computing; 2013. R: A language and environment for statistical computing URL: https://www.r-project.org/ [accessed 2018-03-30] [WebCite Cache ID 6yIXbZNYd]
48. Bates D, Mächler M, Bolker B, Walker S. 2014. Fitting linear mixed-effects models using lme4 URL: https://arxiv.org/abs/1406.5823 [accessed 2018-03-30] [WebCite Cache ID 6yIY4tl1j]
49. Robin X, Turck N, Hainard A, Tiberti N, Lisacek F, Sanchez JC, et al. pROC: an open-source package for R and S+ to analyze and compare ROC curves. BMC Bioinformatics 2011 Mar 17;12:77 [FREE Full text] [doi: 10.1186/1471-2105-12-77] [Medline: 21414208]
50. Akaike H. A new look at the statistical model identification. IEEE Trans Autom Control 1974;19(6):716-723. [doi: 10.1109/TAC.1974.1100705]
51. Venables W, Ripley B. Bagualu. 2003. Modern applied statistics with S URL: http://www.bagualu.net/wordpress/wp-content/uploads/2015/10/Modern_Applied_Statistics_With_S.pdf [accessed 2018-03-30] [WebCite Cache ID 6yIYWOcPo]
52. Hosmer Jr DW, Lemeshow S, Sturdivant RX. Applied Logistic Regression, 3rd Edition. New York, NY: John Wiley & Sons; 2013.
53. Castro AL, Gustafson EL, Ford AE, Edidin JP, Smith DL, Hunter SJ, et al. Psychiatric disorders, high-risk behaviors, and chronicity of episodes among predominantly African American homeless Chicago youth. J Health Care Poor Underserved 2014 Aug;25(3):1201-1216 [FREE Full text] [doi: 10.1353/hpu.2014.0124] [Medline: 25130234]
54. De Rosa CJ, Montgomery S, Hyde J, Iverson E, Kipke M. HIV risk behavior and HIV testing: a comparison of rates and associated factors among homeless and runaway adolescents in two cities. AIDS Educ Prev 2001 Apr;13(2):131-148. [Medline: 11398958]
55. Tucker JS, Ryan GW, Golinelli D, Ewing B, Wenzel SL, Kennedy DP, et al. Substance use and other risk factors for unprotected sex: results from an event-based study of homeless youth. AIDS Behav 2012 Aug;16(6):1699-1707 [FREE Full text] [doi: 10.1007/s10461-011-0017-9] [Medline: 21932093]
56. Owens E, Behun RJ, Manning JC, Reid RC. The impact of Internet pornography on adolescents: a review of the research. Sex Addict Compulsivity 2012;19(1-2):99-122. [doi: 10.1080/10720162.2012.660431]
57. Braun-Courville DK, Rojas M. Exposure to sexually explicit Web sites and adolescent sexual attitudes and behaviors. J Adolesc Health 2009 Aug;45(2):156-162. [doi: 10.1016/j.jadohealth.2008.12.004] [Medline: 19628142]
58. Santa Maria D, Narendorf SC, Ha Y, Bezette-Flores N. Exploring contextual factors of youth homelessness and sexual risk behaviors: a qualitative study. Perspect Sex Reprod Health 2015 Dec;47(4):195-201. [doi: 10.1363/47e6715] [Medline: 26575948]
59. Narendorf SC, Cross MB, Santa Maria D, Swank PR, Bordnick PS. Relations between mental health diagnoses, mental health treatment, and substance use in homeless youth. Drug Alcohol Depend 2017 Dec 01;175:1-8. [doi: 10.1016/j.drugalcdep.2017.01.028] [Medline: 28364629]
60. Santa Maria D, Narendorf SC, Bezette-Flores N, Ha Y. “Then you fall off”: youth experiences and responses to transitioning to homelessness. JFS 2015;15(1) Article 5.

Abbreviations

ADHD: attention deficit-hyperactivity disorder
AIC: Akaike information criterion
EMA: ecological momentary assessment
GLMM: generalized linear mixed models

LGBT: lesbian, gay, bisexual, and transgender
OR: odds ratio
PTSD: posttraumatic stress disorder
ROC: receiver operating characteristic curve
STI: sexually transmitted infection

Edited by G Eysenbach; submitted 21.09.17; peer-reviewed by R Paz Castro, T Wray; comments to author 09.12.17; revised version received 02.02.18; accepted 02.02.18; published 10.04.18.

Please cite as:
Santa Maria D, Padhye N, Yang Y, Gallardo K, Businelle M
Predicting Sexual Behaviors Among Homeless Young Adults: Ecological Momentary Assessment Study
JMIR Public Health Surveill 2018;4(2):e39
URL: http://publichealth.jmir.org/2018/2/e39/
doi:10.2196/publichealth.9020
PMID:29636318

Original Paper

Developing Sustainable and Impactful Mobile Phone HIV Testing Interventions for Spanish-Speaking Men Who Have Sex With Men in the United States: Lessons Learned From Informative Interviews

Jason Mitchell[1], MPH, PhD; Maria Beatriz Torres[2], PhD; Lucy Asmar[3], MPH; Thu Danh[4], MPH; Keith J Horvath[4], PhD

[1]Office of Public Health Studies, University of Hawai'i at Mānoa, Honolulu, HI, United States

[2]Department of Communication Studies, Gustavus Adolphus College, St Peter, MN, United States

[3]Department of Public Health Sciences, University of Miami Miller School of Medicine, Miami, FL, United States

[4]Division of Epidemiology and Community Health, School of Public Health, University of Minnesota, Minneapolis, MN, United States

Corresponding Author:
Jason Mitchell, MPH, PhD
Office of Public Health Studies
University of Hawai'i at Mānoa
1960 East-West Road
Biomed T110
Honolulu, HI, 96822
United States
Phone: 1 808 956 3342
Fax: 1 808 956 3342
Email: jasonmit@hawaii.edu

Abstract

Background: Although many men who have sex with men (MSM) test for HIV at least once in their lifetime, opportunities to improve regular HIV testing, particularly among Hispanic or Latino MSM, is needed. Many mHealth interventions in development, including the ones on HIV testing, have primarily focused on English-speaking white, black, and MSM of other races. To date, no studies have assessed app use, attitudes, and motivations for downloading and sustaining use of mobile apps and preferences with respect to HIV prevention among Spanish-speaking, Hispanic MSM in the United States.

Objective: The primary aims of this study were to determine what features and functions of smartphone apps do Hispanic, Spanish-speaking MSM believe are associated with downloading apps to their smartphones, (2) what features and functions of smartphone apps are most likely to influence men's sustained use of apps over time, and (3) what features and functions do men prefer in a smartphone app aimed to promote regular testing for HIV.

Methods: Interviews (N=15) were conducted with a racially diverse group of sexually active, HIV-negative, Spanish-speaking, Hispanic MSM in Miami, Florida. Interviews were digitally recorded, transcribed verbatim, translated back to English, and de-identified for analysis. A constant-comparison method (ie, grounded theory coding) was employed to examine themes that emerged from the interviews.

Results: Personal interest was the primary reason associated with whether men downloaded an app. Keeping personal information secure, cost, influence by peers and posted reviews, ease of use, and functionality affected whether they downloaded and used the app over time. Men also reported that entertainment value and frequency of updates influenced whether they kept and continued to use an app over time. There were 4 reasons why participants chose to delete an app—dislike, lack of use, cost, and lack of memory or space. Participants also shared their preferences for an app to encourage regular HIV testing by providing feedback on test reminders, tailored testing interval recommendations, HIV test locator, and monitoring of personal sexual behaviors.

Conclusions: The features and functions of mobile apps that Spanish-speaking MSM in this study believed were associated with downloading and/or sustained engagement of an app generally reflected the priorities mentioned in an earlier study with English-speaking MSM. Unlike the earlier study, Spanish-speaking MSM prioritized personal interest in a mobile app and de-emphasized the efficiency of an app to make their lives easier in their decision to download an app to their mobile device. Tailoring mobile apps to the language and needs of Spanish-speaking MSM is critical to help increase their willingness to download a mobile app. Despite the growing number of HIV-prevention apps in development, few are tailored to Spanish-speaking MSM, representing an important gap that should be addressed in future research.

(JMIR Public Health Surveill 2018;4(2):e45) doi:10.2196/publichealth.8992

KEYWORDS

smartphone; HIV testing; HIV prevention; men who have sex with men; Spanish

Introduction

Background

HIV testing is paramount to primary prevention and serves as the gateway into medical care upon receiving a seropositive diagnosis. Compared with other groups in the United States, HIV disproportionately affects gay, bisexual, and other men who have sex with men (MSM). Although the number of new HIV infected results from 2010 through 2014 remained stable among MSM, MSM accounted for 82% of new HIV diagnoses among males and 68% of total new diagnoses in 2015 [1]. If current HIV rates continue, the Centers for Disease Control and Prevention estimates that 1 in 6 MSM will be diagnosed with HIV in their lifetime, which differs by race: 1 in 2 for blacks, 1 in 4 for Latinos, and 1 in 11 for white MSM [2]. Among Latino MSM, diagnoses of HIV increased by 13% from 2010 to 2014, and Latino MSM in the Southern region are more likely to acquire HIV from male-to-male sexual contact than from other routes of transmission in other regions of the United States (eg, primarily intravenous drug use if in the Northeast region) [3]. This body of epidemiological evidence calls for the need to improve HIV prevention efforts for Latino MSM, including testing for HIV and sexually transmitted infections.

Although many MSM test for HIV at least once in their lifetime, opportunities to improve regular HIV testing among Hispanic or Latino MSM are evident. Findings from the 2008 HIV Behavioral Surveillance System found that 63% of Hispanic or Latino MSM had been tested for HIV in the past 12 months, and that recent testing (within the past year) was strongly associated with being of younger age, having seen a health care provider, and disclosure of male-to-male attraction or sexual behavior to a health care provider [4]. Joseph and colleagues reported that among the 608 Hispanic or Latino MSM residing in New York City or Miami, Florida, 31% were test avoiders (those who had never been tested or had last gotten tested for HIV 5 years ago), and 69% received their last HIV test more than 1 year ago. Hispanic or Latino MSM test avoiders were less likely to have incomes over US $10,000 and to have seen a health care provider in the past year [5]. Moreover, in a study with 538 Hispanic or Latino MSM residing in South Florida, Fernandez et al found that 70.4% of men were repeat testers, defined as having been tested for HIV 3 times in their lifetime; men who were repeat testers were more likely to be older, more educated, having a history of diagnosis of sexually transmitted disease, and a greater number of sex partners [6]. Regular testers, defined by receiving 2 tests per year for minimum of 1 year, accounted for 43% of men in the sample; regular testers were more likely to be younger and have lower risk perceptions of acquiring HIV [6]. The aforementioned studies highlight that certain demographics were associated with either being a repeat tester for HIV or an HIV test avoider among Hispanic or Latino MSM in the United States.

One promising area that has garnished increased attention in HIV prevention is mHealth. The World Health Organization defines mHealth as “medical and public health practice supported by mobile devices, such as mobile phones, smartphones, patient monitoring services, personal digital assistants, and other wireless devices” [7].

Smartphones, which include global positioning systems (GPS), text messaging (short message service, SMS), and app features, have become almost ubiquitous among Hispanic adults, including MSM, as they tend to be early adopters of newer technologies [8]. Preliminary work, much of which is qualitative, has begun to understand how best to optimize content, design, and delivery features of mHealth technologies for HIV prevention and testing [9-12]. In the study by Goldenberg and colleagues, MSM preferred their HIV prevention app to (1) have an educational component to guide their decisions for which test is best for them and prevention options, (2) be interactive and engaging with personalized feedback about their own sexual behaviors, (3) provide a social-networking component with other MSM, (4) use language that is simple and understandable to the community, and (5) address privacy concerns by ensuring that the app is from a credible source and having secure messaging features [10]. In the Schnall and colleagues’ study, high-risk MSM described wanting a self-information management system, focus on staying healthy, HIV testing, a chat or communication function, and other related resources in an HIV-prevention app [12]. High-risk MSM in the study by Aliabadi and colleagues stated that they preferred HIV prevention apps to include key information about HIV testing (eg, options) and support group information, motivational messages about addressing sexual encounters in which men intend to use condoms but do not, behavioral skills about negotiating safer sex, and understanding signs of HIV infection [9].

In the study by Mitchell et al, focus groups were conducted with English-speaking MSM (N=34, 68% white, 27% Hispanic) from Minneapolis, Minnesota, and Miami, Florida, to better understand what features and functions of a mobile app they believe are related to whether they download and continue to stay engaged with the app on their mobile device [11]. Men stated that they tend to download apps that they perceive to be low cost (eg, free), secure, and efficient in helping them save time and for adding convenience to their lives [11]. Features and functions that increased men’s willingness to remain engaged with apps on their mobile device over time (some of which were also related to downloading the app initially) included how useful and necessary the app was to them, reviews from other users and recommendations from friends, ease of use, how reliable the app was, and how frequently the app was updated [11]. When asked about their preferences for an HIV-testing app, men stated that they would like an app that contained a tailored (to their sexual activity) testing interval recommendation, an HIV-test locator, and a way to monitor their personal sexual behaviors [11].

Objective

The studies described above provide a growing body of literature to inform new mHealth interventions; however, these studies primarily focus on English-speaking white, black, and MSM of other races. To our knowledge, no studies have asked primarily Spanish-speaking MSM in the United States about their app use, attitudes, and motivations for downloading and sustaining use of mobile apps, and preferences with respect to HIV prevention. Building on previously published work (eg, [11]), we conducted individual interviews with Spanish-speaking, HIV-negative MSM to explore what features and functions of mobile apps that participants believed were related to downloading and continuing to use apps on their mobile device, which could then be applied toward development of an app to promoting regular HIV testing. We recruited Hispanic, Spanish-speaking MSM living in Miami, Florida, to assess whether findings from previous work align with this group of men living in a different area of the United States. The primary objectives of this study were to answer the following questions:

1. What features and functions of smartphone apps do Hispanic, Spanish-speaking MSM believe are associated with downloading apps to their smartphones?
2. What features and functions of smartphone apps are most likely to influence mens' sustained use of apps over time?
3. What features and functions do men prefer in a smartphone app aimed to promote regular testing for HIV?

Methods

Recruitment and Eligibility

Recruitment for the study sample was conducted through targeted Facebook advertising, flyers displayed in places where MSM frequent, and through referrals as a form of snowball sampling.

Facebook advertisements were conducted in early 2016 and targeted individuals who described themselves as Spanish or bilingual (Spanish- and English- speaking) males who were aged at least 18 years, living in the metro area of Miami, Florida, and interested in men. Each targeted advertisement contained a picture of a Hispanic male with a brief study description that stated participation included a confidential, individual interview with a link to the eligibility screener. Each of the 2 campaigns lasted 3 days with advertisements appearing on the home page, desktop newsfeed, and mobile newsfeed of persons whose Facebook profiles met the targeting criteria.

Study flyers were displayed at establishments that MSM frequent (eg, bars and nightclubs) and at community-based organizations and health clinics that provide HIV testing in Miami, Florida metro area. Flyers contained a brief description of the study with a weblink for the eligibility screener. In addition, participants who were eligible and enrolled in the study were asked to share information about the study with their Spanish-speaking MSM peers.

The targeted Facebook advertisements generated 2131 Web clicks to the eligibility screener, which resulted in 106 individuals completing the eligibility screener. Out of the 15 participants (14.2%) that were eligible and provided consent to participate in the study, 5 learned about the study from Facebook advertisements, 9 were referred by other participants, and 1 participant was recruited via a flyer.

To participate, individuals had to have met the following self-reported eligibility criteria: (1) identify as male, (2) be 18 years of age or older, (3) live in the metro area of Miami, Florida, (4) be HIV-negative or have an unknown serostatus, (5) be a Spanish or bilingual (Spanish and English) speaker, (6) have had anal sex with another man within the past year, and (7) be a current owner of a smartphone. Once deemed eligible via the online screener, participants were asked to provide consent electronically and provide 2 contact methods so that the research team could coordinate and schedule their interviews.

Interviews

After providing consent and contact information, participants were contacted by the study team to schedule an appointment for their one-time, in-person, confidential interview. All interviews were conducted in Spanish, audio recorded, transcribed verbatim in Spanish, translated to English, checked for accuracy, and then de-identified for anonymity purposes. The interview was semistructured, focusing on Spanish-speaking MSM's smartphone preferences (eg, likes, dislikes), apps they currently have and use on their smartphone, and their reasons for downloading and keeping apps on their smartphone over time—defined as 3 months or longer. In addition, participants were asked their opinions about health-related apps and suggestions to design a smartphone app for MSM to encourage regular testing for HIV. Finally, participants were asked to provide feedback on potential features of an HIV-testing app, such as informational messages, testing interval developer, monitoring their sexual activities, and a GPS function to locate nearby HIV-testing locations. Upon completion of the interview, each participant received an incentive of US $50 for his time.

Data Analysis

Content analysis and grounded theory were used to analyze the data. A team of 2 independent research associates coded data from the interviews. One of the independent research associates was formally trained in qualitative methods and had several years of experience as a qualitative researcher. This individual led several meetings to (1) train the other associate in qualitative coding before the coding process began and (2) to facilitate discussions and resolve disagreements regarding coding differences, after each associate coded the interviews.

Both inductive [13] and deductive processes were used as each associate independently coded the interview transcripts and created a codebook in a 3-step process:

1. Each associate independently read all interviews and identified a preliminary set of codes or items found in the

interviews (an inductive process). These emerging codes were statements or words that referred to events, behaviors, or activities that occurred frequently or rarely, and which were similar and/or different to other codes in the transcript.

2. Following the grounded theory approach, each associate independently clustered or grouped identified codes or items that fit together into larger categories of meaning (themes or patterns) and those who did not fit any theme (a deductive process) [14]. In doing so, associates paid specific attention to the contexts and behaviors of the phenomena under study as emerging from interviewees' statements.
3. With this initial-coding scheme (or codebook), each associate recoded the data using the identified themes or patterns. During this process, each associate refined, added, and/or eliminated codes and themes to ensure all possible meanings were accurately captured. A spreadsheet was used to organize codes under each pattern or theme and all quotes associated with each code.

The 2 associates met to review the coding spreadsheets (codebooks) that emerged from the individual process mentioned above. They first looked at agreements. Agreement implied that both associates had identified the same codes or themes or patterns from the same interview quotes. Then, disagreements were discussed. Disagreements ranged from using different labels to name a code or a pattern, to finding that one coder identified a code or a pattern but the other associate did not. In those cases, both associates collectively made a decision as to how the disagreement would be resolved. In those few instances, associates collectively modified codes where inconsistencies were found and ensured all possible meanings were fully and accurately captured. Finally, a master Excel file with an identical codebook was created to reflect the agreements reached by both research associates, which they both then used and applied across all interviews.

Results

Characteristics of the Sample

Table 1 provides sociodemographic characteristics of the study sample. On average, men were 32.4 years old; their ages ranged from 18 to 68 years. All men ethnically self-identified as Hispanic; with respect to race, 10 of the men were self-reported as white, 1 as black, and 4 as another race. Regarding HIV testing, 7 men tested 1 to 3 months before study enrollment and 2 other men had gotten tested 4 to 6 months before the study. The timeframe of when the remaining participants had last gotten tested for HIV varied; 2 men reported that they had never been tested for HIV. Table 2 provides information about the themes and associated definitions for participants' reasons to download and continue the use of apps on their smartphone, as well as their preferences for an HIV testing app.

Table 1. Sociodemographic characteristics of the study sample.

Characteristic	Value
Total number of participants, N	15
Age in years, mean (range)	32.4 (20-68)
Race, n (%)	
White	10 (67)
Other	4 (27)
Black	1 (7)
Ethnicity, n (%)	
Hispanic	15 (100)
Most recent HIV test, n (%)	
1-3 months ago	7 (47)
4-6 months ago	2 (13)
7-9 months ago	1 (7)
10-12 months ago	1 (7)
More than 1 year ago	1 (7)
5 or more years ago	1 (7)
Never been tested	2 (13)
Type of smartphone owned, n (%)	
iOS	8 (53)
Android	6 (40)
Windows	1 (7)

Table 2. Themes, definitions, and participants' general reasons to download and continue the use of apps and their preferences for an HIV testing app.

Category	Theme	Definition
Reasons for downloading an app	Personal interest	Downloading the app because it is of personal interest.
Reasons for downloading and using the app over time	Keeping personal information secure	The app being secure in terms of access to and/or protecting personal information.
	Cost	Download an app if it requires an initial payment or keep using an app if it requires additional payments.
	Influence by peers and posted reviews	Download and/or continue to use a certain app because of positive reviews by peers, friends, or app ratings.
	Ease of use	Downloading and sustaining use of an app because of the ease of use.
	Functionality	Discuss the importance of the app being functional and fulfilling a specific need.
Reasons for keeping and using the app over time	Entertainment	Discussed importance of enjoying the app to continue its use.
	Updates	Frequency in which an app is updated.
Preferences for HIV testing app features and functionality	Informational messages	Discussed opinions about receiving informational messages about the importance of testing for HIV and other sexually transmitted diseases.
	Monitor sexual activity	Opinions shared about monitoring their own sexual behaviors.
	Recommended testing intervals with dates	Discussed receiving personalized, recommended testing intervals with specific dates of when to be tested next.
	Details about testing locations and HIV test locator	Opinions shared about wanting to know nearby locations to test and information about the testing sites.

Reasons to Download an App

Personal interest in an app was the theme discussed as a reason for downloading an app. More than a quarter of participants shared that the app should be of personal interest in order to download it:

> *If it is of my interest I download it.* [49 years old, white, last HIV test 1-3 months before study]
>
> *It has to be something that interests me. Anything, but that it interests me.* [38 years old, other race, last HIV test was 5 or more years before the study]

Reasons to Download and Use the App Over Time

Interviews offered several reasons for downloading and using the app over time (ie, 3 months or longer), such as security of personal information, the cost of the app, influence from reviews and peers, the ease of use, and its functionality.

Keeping Personal Information Secure

All participants agreed that to download an app, it must be secure at keeping their personal information confidential and private. As this participant aptly stated:

> *...that's 100% important, because if it is not safe, what sense does it make to download it?* [42 years old, black, last HIV test was 1 to 3 months before the study]

Men also perceived security to be an important factor that influenced them to download and use an app over time, particularly with respect to the types of data they were comfortable sharing with the app:

> *I care a lot about that...only hotels and airlines and that's it. The rest I do not put my credit card or my personal information anywhere.* [38 years old, other race, last HIV test 5 or more years before the study]

Cost

The majority of the participants (n=12) stated that they were willing to pay to download an app or pay a monthly subscription fee as long as the app is useful to them:

> *It depends on the application. For example, if it is necessary to use, I cannot live without it, of course. But for example, I would not pay to download a social network on my mobile phone.* [21 years old, white, last HIV test was 1 to 3 months before the study]

Another participant stated:

> *Depending on how good is the application and how useful it is; I consider on paying monthly subscription.* [23 years old, Other, never been tested for HIV]

Fewer men (n=3) expressed that they were not willing to pay to use an app and chose to only download free apps:

> *Most of the time they don't have a cost, because you look for free applications...I never, when they start to ask information about a card I leave it there.* [49 years old, white, last HIV test was 1 to 3 months before the study]

Another interviewee stated a free trial is a must:

> *I think it's the best way. I think the best way because one really sees if it works or does not work...it happened to me that I downloaded applications, and*

> *then paid five dollars and it does not work.* [30 years old, white, last HIV test was 1 to 3 months before the study]

Influence by Peers and Posted Reviews

Participants were influenced to download an app by the reviews that came from friends or word of mouth:

> *That is, if all my friends have and say, "Hey download it, is good," obviously I am inclined to download* [27 years old, other race, last HIV test was 1 to 3 months before the study]

Other participants were influenced by the app ratings:

> *See the reviews, reports that people have about the application and if it is good or bad and all that.* [23 years old, other race, never been tested for HIV]

Few participants mentioned their influence to download an app came from the media or social media:

> *Well, if I see an application on TV.* [49 years old, white, last HIV test was 1 to 3 months before the study]

Five participants identified that friends using the app was important for them to keep using the app over time:

> *If they have it yes and if it is good...* [35 years old, other race, last HIV test was 7 to 9 months before the study]

Other 3 participants partially agree with the statement above:

> *Yes, medium, if my friends leave the application, but I still made me useful, I keep it.* [23 years old, other race, never been tested for HIV]

Six participants claimed that friends using the app had no influence in them keeping an app:

> *No, if the need does not matter that they do not use it and I do. It does not matter. It is not a priority for me.* [42 years old, black, last HIV test was 1 to 3 months before the study]

Ease of Use

Besides security and cost, the majority of interviewees identified the importance of intuitive features and functions to download an app:

> *If I download it [app], it has a friendly interface and is not very difficult to use.* [30 years old, white, last HIV test was 1 to 3 months before the study]

Another participant stated:

> *...and one that is also easy to handle, because if it is too complicated, I do not care. So basically, that works for me and easy to navigate and use.* [42 years old, white, last HIV test was 1 to 3 months before the study]

Ease of use also influenced men on whether they kept using the app over time:

> *...easy to navigate within the application or not. If it is too complicated, like it doesn't go...* [20 years old, white, never been tested before]

Another participant said:

> *Very important because...I handle the Smartphone to its fullest, but for others it becomes complicated perhaps...you have to have an easy interface.* [30 years old, white, last HIV test was 1 to 3 months before the study]

Functionality

Participants also indicated that the app needed to be functional and to help fill a particular need of theirs in order for them to download it and keep using an app over time. This participant aptly expressed the importance that function and perceived need for the app influence their desire to download it:

> *First the functionality of the application, if it's really going to work for me, for what I want, for what I need.* [30 years old, white, last HIV test was 1 to 3 months before the study]

Another man in the group expressed how functionality of the app promotes continued use of the app over time:

> *The main reason is it useful to me, as I had said earlier, the utility that has the application and I would say would be the main reason to keep an application that is useful to me. The usefulness of the application.* [42 years old, white, last HIV test was 1 to 3 months before the study]

Reasons for Using the App Over Time

Themes solely related to mens' reasons for using an app over time (ie, 3 months or longer) included entertainment and app updates.

Entertainment

More than half of the interviewees (8) explained that they used an app over time if it provided them with some form of entertainment:

> *The truth is that if keeps me entertained or I get notifications, such as news and all that, then yes, if not, I delete it.* [23 years old, other race, never been tested for HIV]

Another participant stated:

> *If entertains me, if I find useful, if I finished- is, I don't know. That is, if I end up liking, obviously I will continue using and all that. So I think that, if entertains me.* [20 years old, white, never been tested for HIV]

Updates

In addition to entertainment, only 3 interviewees mentioned the importance of frequent updates for using the app over time:

> *If there are constant updates, all that. I find it important that they are constantly updated, to do new things, new features, all that.* [20 years old, white, never been tested for HIV]

Another participant said:

> *Do not get stuck much. That they are always doing updates...you have something new.* [20 years old, white, last HIV test was 10 to 12 months before the study]

Reasons for Deleting an App

There were 4 reasons why participants chose to delete an app—dislike, lack of use, cost, and lack of memory or space. Several participants mentioned dislike (n=4) and lack of use (n=4) as major reasons for deleting an app:

> *It depends. That is, if it is something of continuous necessity, or continuous use, I keep using it...and if I no longer need it, I delete it, so it does not take up space on the phone.* [27 years old, other race, last HIV test was 1 to 3 months before the study]

Another participant stated:

> *I try it first, if I don't like it maybe in the moment I de-install...two or three applications per month that I didn't like, I discard them.* [49 years old, white, last HIV test was 1 to 3 months before the study]

Likewise, a few interviewees (n=4) stated cost and lack of memory as reasons for deleting an app:

> *I eliminate them...if they ask for a price I can't afford.* [68 years old, white, last HIV test was 1 to 3 months before the study]

Another participant said:

> *Every time I run out of memory, I decide to delete applications that have not use in a long time.* [21 years old, white, last HIV test was 1 to 3 months before the study]

Preferences for HIV Testing App Features and Functionality

The final part of the interview explored participant's attitudes toward potential features of a hypothetical HIV-testing app. The specific features explored included, informational messages, sexual activity monitor, recommended testing interval with dates, details about testing locations, and an HIV-testing locator.

Informational Messages

The majority of interviewees would like to receive HIV testing interval messages. The discussion of the informational messages centered around 2 primary components—frequency and format. For frequency, some participants discussed that they would like to receive informational messages once a week:

> *Well, weekly...It's fine, because that is, as I said before helps you keep in touch with yourself, so to speak. Your health is also important because it is not - is not a myth that health always left aside.* [25 years old, white, last HIV test was 4 to 6 months before the study]

Others would like extended time intervals between these messages:

> *Once every two weeks.* [27 years old, other race, last HIV test was 1 to 3 months before the study]

Another participant stated:

> *Once a month for you to be aware of how you can take care of yourself...* [35 years old, other race, last HIV test was 7 to 9 months before the study]

For format, some participants would rather receive the message through the app:

> *It can be like an alert if you have the application, yes, like a notification. I do not - in my case it would have no problem. But a notification seems very well to me.* [25 years old, white, last HIV test more than a year before the study]

Others would rather receive the informational message via email or text:

> *I think through text is cool, super cool. Because you know there are people who sometimes do not like notifications that comes up on their screen...And I think by text or email, for me it would be fine.* [20 years old, white, last HIV test was 10 to 12 months before the study]

Sexual Activity Monitor

In addition to receiving test interval messages, most participants agreed with the idea of including a sexual monitor feature in the app:

> *I think it's fun that you can realize: "Oh, this is the amount of times I've had sex this week." "This is the amount of times I've drink-." Then, like: "Can I do better?...That is: "Can I stop? Can I take more care of myself?"* [21 years old, white, last HIV test was 4 to 6 months before the study]

Another participant stated:

> *Yes, that is as the application has to do with that part of what- to make checkups regularly- I think it's okay to keep [track of] your activity, therefore, that, you know, if you had relations with six people this week is like well, you know you have to make a checkup.* [20 years old, white, never been tested before]

In contrast, one participant indicated that he did not see the purpose of this sexual activity monitor:

> *I do not see the point. I mean, why would someone do something like that? I mean, I do not see what could be the purpose of that. I mean, I would not see the usefulness.* [38 years old, other race, last HIV test 5 or more years before the study]

Recommended Testing Intervals With Dates

Although 3 men did not provide an opinion, all other participants universally liked the recommended testing interval as an app feature. Many of the participants succinctly shared their support for this idea by saying:

> *It sounds good. Sounds good to me.* [25 years old, white, last HIV test more than a year before the study]

In addition to supporting and liking this idea, other participants provided more context as to why this idea of being provided with recommended HIV-testing intervals with dates appealed to them:

> *In the event that someone will become interested, I think is super good, because if you are giving it as some guidance. That's what you were talking about a guide, it is very clear. If that can send with a hidden notification, great.* [38 years old, other race, last HIV test 5 or more years before the study]

Details About Testing Locations and HIV-Testing Locator

Another feature with unanimous consensus among participants was the inclusion of GPS function that would allow them to find the nearest testing center should be included in the app:

> *Very useful, because there is often the laziness factor and time factor as to the availability of people, not often to use them. And if you have something that reports what is closest to you and that will save time, you will save money, then so be it.* [25 years old, white, last HIV test was 4 to 6 months before the study]

And another participant stated:

> *Fabulous, because see I was about to do it a while ago and I couldn't find a place.* [49 years old, white, last HIV test was 1 to 3 months before the study]

Discussion

Principal Findings

A growing body of literature has examined MSM's preferences for app design, features, and functionality to direct the development of mobile apps addressing a variety of HIV prevention outcomes (eg, HIV testing, reduce engagement in condomless anal sex) [10,15]. However, to our knowledge, this is the first study to explore men's preferences in a sample of Spanish-speaking MSM, who are important targets for HIV prevention services [2]. Overall, this study showed that Spanish-speaking MSM reported many of the same considerations (eg, ease of use, perceived usefulness, security) that impacted their adoption and use of mobile apps and other technologies as shown in prior studies of primarily English-speaking MSM [11,15]. In addition, Spanish-speaking MSM noted that having a strong personal interest in the content of the mobile app strengthened their motivation to download the app on their phone. These results are discussed in more detail below.

Prior work has shown that English-speaking MSM in Minneapolis, Minnesota, and Miami, Florida, who participated in focus groups considered cost, security, and efficiencies afforded by the app in their decision to download it, whereas influence by others (either through reviews or friends) and the app's usefulness was related to both downloading and sustaining use (Table 3) [11]. Continuing to use apps on their mobile device was primarily attributed to how easy the app was to use, the reliability of the app, and how frequently it was updated. Compared with these findings, Spanish-speaking MSM also endorsed the importance of these factors but believed that many of them (ie, cost, security, influence by reviews or friends, usefulness, and ease of use) contributed to their willingness to both download the app and continue using it (Table 3). Spanish-speaking MSM did not identify an app's ability to make their life more efficient as a primary motivator to downloading apps to their mobile device; however, they did state that a strong personal interest in an app was important to motivate them to download the app. This finding brings to light the importance of developing tailored HIV prevention apps for Spanish-speaking MSM that address their unique values and preferences (both their language preferences and preferences for content and features) in order to increase their personal interest in the app and motivate them to download it to their phone.

A number of in-person [16-19] and computer-mediated [20] intervention approaches have been tailored to the needs of Spanish-speaking MSM. However, to our knowledge, no Spanish-language HIV prevention mobile apps have been developed and tested for Spanish-speaking MSM in the United States. These finding suggest that this may be an important avenue for future research on using mobile apps to deliver HIV prevention interventions to this group.

Men in this study were also asked to reflect on their preferences for features of an HIV-testing mobile app. There was variation among men in this sample for the frequency and delivery method for informational messages (eg, about risk of HIV or the importance of HIV testing). Some men wanted informational messages to be sent through the app (ie, as a notification), whereas other men preferred an SMS (text) message. Some men preferred weekly informational messages, whereas other men expressed the desire for less frequent messages. Together, this suggests that giving Spanish-speaking MSM options for frequency and format of informational messages may be the most effective way to accommodate their varied preferences for these factors.

Table 3. Comparison of themes related to downloading, downloading and sustained use, and sustained use of smartphone apps between English-speaking men who have sex with men (Mitchell et al, 2016) and Spanish-speaking men who have sex with men (MSM).

Reason	Downloading		Downloading and sustained use		Sustained use	
	English	Spanish	English	Spanish	English	Spanish
Cost	✓			✓		
Security	✓			✓		
Efficiency	✓					
Personal interest		✓				
Influence from reviews or friends			✓	✓		
Usefulness or functionality			✓	✓		
Ease of use				✓	✓	
Reliability					✓	
Updates					✓	✓
Entertainment						✓

In contrast, there was nearly universal or universal support for mobile app features that included an HIV-test locator (100% support), a recommended HIV test interval period for men to get tested (80% support), and to monitor their sexual activity (93% support). English-speaking MSM in a prior study also unanimously supported an HIV-test locator and a recommendation for the best date range to be tested for HIV; however, approximately three-fourths supported a feature for monitoring their sexual behaviors [11]. Differences in support for this feature may be due to different beliefs about the usefulness of sexual activity monitoring, or concerns about the security of storing sexual data on their mobile device. Both Spanish- and English-speaking MSM stated that security considerations were important to them [11]; perhaps suggesting the perceived usefulness of this feature may be the primary reason to explain differences in their support for sexual activity monitoring. A separate study also reported some support among English-speaking MSM residing in Atlanta, Seattle, and rural areas of the United States for monitoring sexual activity and feedback on high-risk encounters [10]. Clearly, continued research is needed to more fully understand MSM's comfort level with sexual self-monitoring and how best to include this feature for men who may be culturally unique.

Limitations

This study has important limitations to acknowledge. First, the results from these interviews are not intended to be generalizable to all Spanish-speaking, HIV-negative MSM who are smartphone owners in the United States. Participants in this study resided in one large urban area (Miami, Florida) and, as such, app preferences may differ from Spanish-speaking MSM residing in different regions of the United States. Future research should assess smartphone app preferences among a more geographically diverse sample of Spanish-speaking MSM to determine whether similar attitudes are expressed. Future studies should also consider whether general smartphone app preferences and those related to HIV prevention and testing differ between bilingual MSM and monolingual Spanish-speaking MSM. Second, recall bias may have affected participants' abilities to accurately identify features and functions that influenced them to download and use apps over time. The qualitative, cross-sectional nature of this study prohibits any causal inference, and longitudinal studies are needed to confirm the findings of this study. These results are meant to be first steps in more fully understanding the needs of Spanish-speaking HIV-negative MSM with respect to mobile app preferences for HIV prevention.

Conclusions

This study presents findings of formative work to understand which features and functions of mobile apps that Spanish-speaking HIV-negative MSM perceived as important in their consideration to download mobile apps to their phone and continue to use them over time. Men also provided critical feedback on features of an HIV-testing app they believed are important to incorporate in a future mobile app. Overall, Spanish-speaking MSM endorsed many of the same priorities expressed among English-speaking men in prior studies [10,11,15]. However, men in this study showed high support for monitoring their sexual activities using a mobile app and believed that personal interest was a strong influence on their willingness to download an app to their mobile device. These findings lend support to developing tailored HIV prevention apps for Spanish-speaking MSM, which is a notable gap in current prevention efforts. These findings may help guide future efforts in developing an HIV prevention app for Spanish-speaking MSM, although more research is needed to fully understand how to best frame content in the app and how to reach Spanish-speaking MSM with mobile app interventions. Furthermore, other research with larger samples of Spanish-speaking MSM is warranted to fully address these gaps in knowledge and whether cultural appropriateness plays a role in future development of HIV prevention apps, including those aimed to promote regular HIV testing.

Acknowledgments

This study was funded by the National Institute of Mental Health (R34MH105202).

Conflicts of Interest

None declared.

References

1. Centers for Disease Control and Prevention. HIV surveillance report, vol 27: Diagnoses of HIV Infection in the United States and Dependent Areas, 2015 URL: https://www.cdc.gov/hiv/pdf/library/reports/surveillance/cdc-hiv-surveillance-report-2015-vol-27.pdf [accessed 2017-09-18] [WebCite Cache ID 6tZltZTvD]
2. Centers for Disease Control and Prevention. 2017. HIV among gay and bisexual men URL: http://www.cdc.gov/hiv/group/msm/index.html [accessed 2017-09-18] [WebCite Cache ID 6tZlcGSTR]
3. Centers for Disease Control and Prevention. 2017. CDC Fact Sheet: HIV among Latinos URL: https://www.cdc.gov/nchhstp/newsroom/docs/factsheets/cdc-hiv-latinos-508.pdf [accessed 2017-09-18] [WebCite Cache ID 6tZlOGfrE]
4. Joseph HA, Belcher L, O'Donnell L, Fernandez MI, Spikes PS, Flores SA. HIV testing among sexually active Hispanic/Latino MSM in Miami-Dade County and New York City: opportunities for increasing acceptance and frequency of testing. Health Promot Pract 2014 Nov;15(6):867-880. [doi: 10.1177/1524839914537493] [Medline: 24920606]
5. Oster AM, Russell K, Wiegand RE, Valverde E, Forrest DW, Cribbin M, NHBS Study Group. HIV infection and testing among Latino men who have sex with men in the United States: the role of location of birth and other social determinants. PLoS One 2013;8(9):e73779 [FREE Full text] [doi: 10.1371/journal.pone.0073779] [Medline: 24147151]
6. Fernández MI, Perrino T, Bowen GS, Royal S, Varga L. Repeat HIV testing among Hispanic men who have sex with men--a sign of risk, prevention, or reassurance? AIDS Educ Prev 2003 Feb;15(1 Suppl A):105-116. [Medline: 12630603]
7. World Health Organization. 2011. New horizons for health through mobile technologies based on the findings of the second global survey on eHealth URL: http://www.who.int/goe/publications/goe_mhealth_web.pdf [accessed 2017-09-18] [WebCite Cache ID 6tZl9RNhi]
8. Smith A. Pew Internet. 2015. The Smartphone Difference URL: http://www.pewinternet.org/2015/04/01/us-smartphone-use-in-2015/ [accessed 2017-09-18] [WebCite Cache ID 6tZkYijMF]
9. Aliabadi N, Carballo-Dieguez A, Bakken S, Rojas M, Brown 3rd W, Carry M, et al. Using the Information-Motivation-Behavioral Skills Model to guide the development of an HIV prevention smartphone application for high-risk MSM. AIDS Educ Prev 2015;27(6):522-537.
10. Goldenberg T, McDougal SJ, Sullivan PS, Stekler JD, Stephenson R. Preferences for a mobile HIV prevention app for men who have sex with men. JMIR Mhealth Uhealth 2014;2(4):e47 [FREE Full text] [doi: 10.2196/mhealth.3745] [Medline: 25355249]
11. Mitchell JW, Torres MB, Joe J, Danh T, Gass B, Horvath KJ. Formative work to develop a tailored HIV testing smartphone app for diverse, at-risk, HIV-negative men who have sex with men: a focus group study. JMIR Mhealth Uhealth 2016 Nov 16;4(4):e128 [FREE Full text] [doi: 10.2196/mhealth.6178] [Medline: 27852558]
12. Schnall R, Rojas M, Bakken S, Brown W, Carballo-Dieguez A, Carry M, et al. A user-centered model for designing consumer mobile health (mHealth) applications (apps). J Biomed Inform 2016 Apr;60:243-251. [doi: 10.1016/j.jbi.2016.02.002] [Medline: 26903153]
13. Dillon DR. Grounded theory and qualitative research. In: The Encyclopedia of Applied Linguistics, 1. Hoboken, NJ, USA: Blackwell Publishing Ltd; 2012.
14. Charmaz K. The search for meanings - grounded theory. In: Smith JA, Harré R, Van Langenhove L, editors. Rethinking Methods in Psychology. London, UK: Sage Publications; 1996:0803977328-0803977349.
15. Schnall R, Higgins T, Brown W, Carballo-Dieguez A, Bakken S. Trust, perceived risk, perceived ease of use and perceived usefulness as factors related to mHealth technology use. Stud Health Technol Inform 2015;216:467-471 [FREE Full text] [Medline: 26262094]
16. Martinez O, Wu E, Levine EC, Muñoz-Laboy M, Fernandez MI, Bass SB, et al. Integration of social, cultural, and biomedical strategies into an existing couple-based behavioral HIV/STI prevention intervention: voices of Latino male couples. PLoS One 2016;11(3):e0152361 [FREE Full text] [doi: 10.1371/journal.pone.0152361] [Medline: 27028873]
17. Melendez RM, Zepeda J, Samaniego R, Chakravarty D, Alaniz G. “La Familia” HIV prevention program: a focus on disclosure and family acceptance for Latino immigrant MSM to the USA. Salud Publica Mex 2013;55 Suppl 4:S491-S497 [FREE Full text] [Medline: 25153189]
18. Rhodes SD, Alonzo J, Mann L, Freeman A, Sun CJ, Garcia M, et al. Enhancement of a locally developed HIV prevention intervention for Hispanic/Latino MSM: a partnership of community-based organizations, a university, and the Centers for Disease Control and Prevention. AIDS Educ Prev 2015;27(4):312-332. [Medline: 26241382]
19. Rhodes SD, Daniel J, Alonzo J, Duck S, Garcia M, Downs M, et al. A systematic community-based participatory approach to refining an evidence-based community-level intervention: the HOLA intervention for Latino men who have sex with men. Health Promot Pract 2013;14(4):607-616. [Medline: 23075504]

20. Young SD, Holloway I, Jaganath D, Rice E, Westmoreland D, Coates T. Project HOPE: online social network changes in an HIV prevention randomized controlled trial for African American and Latino men who have sex with men. Am J Public Health 2014 Sep;104(9):1707-1712. [doi: 10.2105/AJPH.2014.301992] [Medline: 25033137]

Abbreviations

GPS: global positioning systems
MSM: men who have sex with men
SMS: short messaging services

Edited by G Eysenbach; submitted 18.09.17; peer-reviewed by R Schnall, T Goldenberg; comments to author 03.12.17; revised version received 11.01.18; accepted 18.02.18; published 24.04.18.

Please cite as:
Mitchell J, Torres MB, Asmar L, Danh T, Horvath KJ
Developing Sustainable and Impactful Mobile Phone HIV Testing Interventions for Spanish-Speaking Men Who Have Sex With Men in the United States: Lessons Learned From Informative Interviews
JMIR Public Health Surveill 2018;4(2):e45
URL: http://publichealth.jmir.org/2018/2/e45/
doi:10.2196/publichealth.8992
PMID:29691205

Original Paper

A Gamified Smartphone App to Support Engagement in Care and Medication Adherence for HIV-Positive Young Men Who Have Sex With Men (AllyQuest): Development and Pilot Study

Lisa Hightow-Weidman[1,2], MPH, MD; Kathryn Muessig[2], PhD; Kelly Knudtson[1], MPH; Mala Srivatsa[3], BA; Ellena Lawrence[3], BA; Sara LeGrand[4], PhD; Anna Hotten[5], PhD; Sybil Hosek[6], PhD

[1]Institute of Global Health and Infectious Diseases, University of North Carolina, Chapel Hill, NC, United States
[2]Department of Health Behavior, Gillings School of Global Public Health, University of North Carolina, Chapel Hill, NC, United States
[3]Ayogo, Vancouver, BC, Canada
[4]Duke Global Health Institute, Duke University, Durham, NC, United States
[5]Department of Epidemiology and Biostatistics, School of Public Health, University of Illinois at Chicago, Chicago, IL, United States
[6]Department of Psychiatry, Stroger Hospital of Cook County, Chicago, IL, United States

Corresponding Author:
Lisa Hightow-Weidman, MPH, MD
Institute of Global Health and Infectious Diseases
University of North Carolina
Bioinformatics Building, Suite 2154
130 Mason Farm Road
Chapel Hill, NC, 27599-7030
United States
Phone: 1 919 843 0033
Email: lisa_hightow@med.unc.edu

Abstract

Background: HIV disproportionately impacts young men who have sex with men (YMSM) who experience disparities across the HIV care continuum. Addressing antiretroviral therapy (ART) adherence among YMSM is an urgent public health priority. Technology-based interventions—particularly mobile health platforms—can provide tailored adherence interventions and allow YMSM to engage and connect with others.

Objective: The objective of this study was to describe the development of *AllyQuest*, a novel, theoretically-based, smartphone app designed to improve engagement in care and ART adherence and social support among HIV-positive YMSM.

Methods: *AllyQuest* was built on an established platform for patient engagement that embeds social networking and fundamental game mechanics, such as challenges, points, and rewards. A medication tracker provides reminders to promote ART adherence via personalized adherence strategies that are user and context specific; a calendar allows for reflection on adherence over time. After iterative development with input from two youth advisory boards, usability testing was conducted to assess app functionality, comprehension of the educational content, use of intervention features, and overall impressions of app relevance and appeal. A 28-day pilot trial was conducted with 20 HIV+ YMSM to evaluate intervention feasibility and acceptability.

Results: Mean age of participants was 21.8 years (range 19-24), and 95% (19/20) of the participants were nonwhite. The mean time of app use was 158.4 min (SD 114.1), with a range of 13 to 441 min. There was a mean of 21.2 days of use (out of a total possible 28 days). There were 222 posts to the daily discussion social wall. Feasibility and acceptability ratings were high. Overall, participants found the app easy to use and navigate, not intrusive, and had few reported technical issues. Higher levels of app usage were positively correlated with HIV self-management outcomes, and there was a statistically significant ($P<.05$) positive association between the number of days logged into the app and knowledge and confidence in ability to reliably take HIV medications.

Conclusions: *AllyQuest* represents a new, highly scalable solution that is well-suited to meet the specific prevention and care needs of HIV+ YMSM. The development of this intervention is both timely and vital, given the urgency of the ongoing HIV epidemic among YMSM.

(JMIR Public Health Surveill 2018;4(2):e34) doi:10.2196/publichealth.8923

KEYWORDS

YMSM; antiretroviral adherence; smartphone app; gamification; social networking

Introduction

HIV Infection Among Young Men Who Have Sex With Men in the United States

In the United States, men who have sex with men (MSM) experience the highest rates of new HIV diagnoses, with young MSM (YMSM) and MSM of color continuing to be significantly impacted [1]. Although the number of diagnoses of HIV infection among MSM remained stable, from 2010 to 2014, the number of diagnoses among MSM in the age range of 13 to 24 years increased by 6% [2]. National HIV Behavioral Surveillance data on MSM from 20 cities found that among black MSM in the age range of 18 to 24 years tested in 2014, 26% were HIV positive, compared with 3% of white MSM. This disparity in HIV prevalence between black and white MSM increased from 2008 to 2014, especially among YMSM [3]. Young Hispanic or Latino MSM had a 20% increase in numbers of diagnoses of HIV infection from 2010-2014 [2].

HIV also disproportionately impacts YMSM across the HIV care continuum, with disparities in linkage, retention, antiretroviral therapy (ART) adherence, and viral suppression [4,5]. YMSM have documented low levels of ART adherence, impeding their likelihood of achieving viral suppression [6]. One study among 13 sites in the US Adolescent Trials Network found only 7% of diagnosed youth (81% male, 72% black, and 70% gay or bisexual) achieved viral suppression [5], which was substantially lower than the estimated 50% viral suppression for all age groups [4,5]. Alarmingly, a recent study of 991 HIV-infected YMSM (aged 15-26 years) found that 69.4% had a detectable viral load; two-thirds of whom (n=458) reported recent condomless anal sex [7].

Barriers to Adherence Among Youth

Individual-level stated barriers to adherence among youth include forgetting, not feeling like taking medication, and not wanting to be reminded of HIV [8]. Additional factors contributing to suboptimal adherence and viral suppression among youth, including YMSM, include low medication adherence self-efficacy [9], psychological distress (depression and anxiety) [10-13], substance use (alcohol, marijuana, and other drugs) [10,13,14], structural barriers (eg, homelessness and health insurance) [15], low social support [11,16], and HIV-related stigma [10,11,17]. Multiple factors are often present, and these *syndemics* are associated with greater likelihood of nonadherence and detectable viral load in a dose-response nature [10,18] that also shows disproportionate impact on minority MSM [18].

There are only a few published interventions focused on supporting ART adherence among YMSM [19,20]. A 2017 systematic review of interventions along the HIV care continuum identified 117 medication adherence interventions. Only 9 (9/117, 8%) focused on adolescents or youth, and only 2 were specifically designed for MSM [19]. A 2014 review of adherence interventions meeting the Centers for Disease Control and Prevention’s criteria for evidence-based interventions found none that exclusively focused on MSM or youth [21]. Due to the often marginalized and stigmatized status that many YMSM living with HIV endure, there is a need for the development of tailored interventions that account for the unique and challenging circumstances they face.

Tailored interventions have been found to produce higher rates of behavior change and maintenance than nontailored programs in a variety of health domains, including HIV [22-24]. Technology-based interventions—particularly mobile health (mHealth) platforms—can provide tailored adherence interventions and allow YMSM to engage and connect with others [25-27]. mHealth tools also offer the capacity to design and deliver tailored content that best meets the specific HIV management challenges faced by YMSM and each individual end user [25]. The fidelity to intervention delivery provided by mHealth and the market saturation of mobile technology ownership across socioeconomic strata [28,30] also provides a high-impact platform that can be taken to scale across and beyond the United States. Daily mobile phone–based contact is acceptable to youth living with HIV and is associated with improved adherence [29]. In this paper, we describe the development, usability evaluation, and subsequent pilot testing of *AllyQuest*, a tailored mobile phone app to increase engagement in HIV care, ART adherence, and social support.

Methods

Intervention Development (September 2015-May 2016)

AllyQuest is a novel, theoretically-based, mobile phone app intervention designed to improve engagement in care and ART adherence and social support among HIV-positive YMSM (target age: 16-24 years). *AllyQuest* development was guided by evidence-based risk reduction and medication adherence interventions [30-32]; health behavior change theories, including social cognitive theory (SCT) [33,34]; narrative communication (eg, storytelling) [35-38]; and the principles of persuasive technology [39]. *AllyQuest* addresses key principles of SCT, including (1) Observational learning by participating in daily activities, (2) Modeling and vicarious experiences (observing and participating in daily discussions, exploration of narrative “choose-your-own-adventure” stories), (3) Self-efficacy and verbal persuasion from expert sources (multimedia knowledge center and tailored messages), and (4) Reinforcements (virtual rewards and achievements) [33,34-40]. Narrative storytelling relies on the modeling of behaviors by similar others and has been shown to foster self-reflection and influence disease self-management [41-43]. The Fogg Behavior Model (FBM) [39] of persuasive technology informed the development of Ayogo’s Empower Platform, the operating system on which *AllyQuest* was developed. According to the FBM, the principal factors to promote behavior change using technology include triggers, ability, and motivation. The design of *AllyQuest* enhances motivation and skills and provides triggers to encourage positive behaviors. App notifications are *triggers* for healthy behaviors. Regular behavioral self-report prompts serve as additional triggers and help participants establish healthy habits. *Ability* is increased through knowledge and by identifying small steps toward target behavioral goals (eg, understanding side effects and knowing how to fill a prescription). Participants also get tips from others who are dealing with similar issues and through narrative stories within the app that reinforce the consequences of healthy and unhealthy behaviors. App *motivators* include social support, rewards, goal setting, and achievements.

Concepts for *AllyQuest* design and overall "look and feel" were informed by research we conducted with HIV-positive YMSM to understand technology utilization, the barriers and facilitators to ART adherence, and the use of an app to assist with adherence [26,44,45]. Gamification elements incorporated into *AllyQuest* included the ability to "level up," earn and redeem in-app virtual currency, and the ability to unlock app features. Throughout development, we worked closely with two YMSM youth advisory boards, composed of eight HIV-positive YMSM, located in Durham, North Carolina and Chicago, Illinois. Youth advisory board activities included 13 in-person sessions and seven rounds of Web-based surveys.

Usability Testing (June 2016-August 2016)

Usability testing was conducted according to established usability guidelines [46] one-on-one with eight (nonyouth advisory board) HIV-positive YMSM aged 16 to 24 years. Participants were guided through app installation on their personal phones. They were asked to explore the app and complete specified tasks within the app without study staff assistance. While exploring the app, participants were asked to "think aloud" and provide a running commentary of their thought processes while performing the tasks [47]. The concurrent think aloud method was chosen to elicit real-time feedback and emotional responses [46,48,49]. Participants then received a guided tour of all app features. Participants were asked about their initial impressions of the app and completed a posttest survey to assess user experience. Participants were then asked to use the app daily for 1 week to assess ongoing functionality, monitor for any technical issues, ensure content comprehension, evaluate intervention features, and describe overall impressions of app relevance and appeal. To facilitate social connectivity, youth advisory board members also used the app during the usability testing period. A semistructured phone interview was conducted at the end of the 1-week testing period.

Intervention Refinement (August 2016-October 2016)

The research team collated all participant feedback into a usability report that was presented to the technological partner. Usability testing revealed several addressable technical bugs and user experience issues that were resolved. Usability participants also provided specific recommendations for content edits and expansion. The research team addressed as many of these suggestions as possible before pilot launch, and the remaining suggestions were prioritized for the next iteration of development.

Pilot Evaluation (October 2016-January 2017)

A 4-week pilot trial was then conducted with 20 HIV-positive YMSM to evaluate intervention feasibility and acceptability. Participants were recruited from a clinic in Chicago that primarily serves impoverished communities. Participants completed a pretest survey and then staff assisted with app download to participants' phones. After 4 weeks of use, participants completed an online posttest survey and a phone-based qualitative interview.

Pilot Trial Measures

Sociodemographic Items

Sociodemographic items assessed age, race or ethnicity, education, income, homelessness, health insurance, and sexual identity.

Feasibility

Usage data was captured through in-app analytics and included number of times per day or week participants accessed the app and average time spent using app, daily number of activities completed and daily discussion questions answered, content of posts, and number of health-focused daily quests completed.

Acceptability

System Usability Scale (SUS) [50] is a 10-item, 5-point Likert scale of subjective assessments of usability. The SUS provides a global measure of system satisfaction and subscales of usability and learnability. For this trial, 9 of the 10 items were used, as one question was deemed duplicative.

Client Satisfaction Questionnaire-8 (CSQ-8) was used to assess global intervention satisfaction. The CSQ-8 has eight items (quality of app, kind of service received from app, app met needs, recommend app to a friend, amount of help received from app, effectiveness of app for dealing with health problem, overall satisfaction, and willingness to use the app again). These domains are assessed on a 4-point response scale with individually specified anchors. Participant responses are scored from 1 to 4, and thus, the possible total scores range from 8 to 32. Higher scores indicate greater satisfaction. The CSQ-8 has demonstrated high internal consistency across a large number of studies and has been used to evaluate technology-based interventions [51-55].

HIV Self-Management

We developed four questions to assess domains of HIV-specific self-management after versus before the 1-month pilot trial. These included feeling connected to others with HIV, knowledge about HIV, ability to effectively manage HIV, and ability to reliably take ART. All outcomes were measured on a 5-point Likert scale, ranging from 1="Much less" to 5="Much more."

Pilot Trial Analysis

Frequencies and measures of central tendency (means, medians, and SDs) were calculated to describe the sample in terms of sociodemographics and acceptability, feasibility, and 4-week retention. Chi-square tests, *t* tests, and correlations were examined to provide preliminary effect estimates. Exact statistical tests were used where possible to account for the small sample size. The Spearman rank order correlation coefficient was calculated to assess the magnitude of association between app usage and HIV self-management outcomes, where a value of 0 indicates no correlation, and higher values indicate stronger association between variables [56]. The Spearman rank order correlation is a nonparametric alternative to the Pearson correlation [57] based on ranks instead of absolute values and is less sensitive to outliers and nonnormal variable distributions. Statistical analyses were conducted using SAS (SAS Institute Inc) software version 9.4 for Windows.

Results

Intervention Development

Youth advisory board members provided feedback on (1) Intervention structure and format (eg, organization of the intervention, appropriateness and appeal of language and images, and ease of navigation); (2) Intervention content and activities (eg, comprehension, acceptability, and relevance); and (3) Overall app impressions (eg, utility, interest, and enjoyment). Youth advisory board members' feedback informed content development in all app components (Table 1).

Table 1. *AllyQuest* intervention components and scientific rationale.

Feature description	Scientific rationale
Profile page	
Privacy features: **t**hese include avatars, pseudonyms, confidential pin number to open app, app time-out after 5 min of inactivity, and medication tracker that allows participants to choose any name (real or made-up) they want for their medication reminder.	Anonymity and privacy recognized as important for YMSM[a] electronic health. Found in work done by our team and others [58,59].
App progression meter: **v**isual display of current app "level" and in-game currency that is visible to other participants. Participants level up and earn in-game currency based on app use. Redeem currency to unlock narratives and other app features.	Game-based elements (eg, levels and competition) influence intervention engagement and impact [60].
Daily discussion	
Social prompts: (eg, How do you remember your medication?) kick-off daily discussions to foster community, peer sharing, model successful behaviors, and provide reinforcement.	Social support and connection with others are important features for apps for HIV-positive YMSM [61].
Medication tracker	
Medication reminder system: discreet personalized reminders and habit building solutions to promote ART[b] adherence.	Medication reminders improve adherence, but may not be sufficient [62].
Tailored adherence strategies: upon initial set-up, participants enter medication details, including the number of times/day and preferred time of day taken and any food restrictions. The app uses this information to provide suggestions on adherence strategies (eg, Take when I brush my teeth). Participants who are having adherence difficulties will received tailored feedback on new strategies and adherence tips.	Dynamic tailoring and unique feedback based on frequent assessments effectively promotes behavior change for many conditions, including HIV prevention and ART adherence [61,63].
Brain builders	
Daily quest: actionable routine tasks help users set goals and build knowledge or skills.	Rated highly by usability and pilot participants. Gamification increases intervention engagement and impact [62,60].
Brain games: quizzes and interactive exercises help users check knowledge and skill	
Knowledge center	
Multimedia: presentation of information that includes HIV-related, safer-sex, relationships and general health and wellness. Users prompted with a reflection question after each article to apply the material to their lives. Visual shows progress toward completing each section.	Formative work of our team and others has identified that HIV+ YMSM desire information on both HIV-related issues and general health and wellness [64].
Character-based narratives	
"Choose-your-own adventure" narratives feature HIV+ YMSM navigating common situations that impact care engagement and ART adherence (eg, unstable housing, substance use, and disclosure). Play through story paths allows the user to face hard choices that impact health, practice problem solving, and succeed or fail in a safe space.	Narrative communication through role modeling has been identified as facilitating health behavior change [35-38].

[a] YMSM: young men who have sex with men.

[b] ART: antiretroviral therapy.

Usability Testing

Usability testing revealed several addressable technical bugs and user experience issues. Given the time frame and cost considerations with app development, we prioritized bug fixes and user experience issues. Although users had other suggestions for app improvement (eg, making the daily discussion more like Facebook and using different imagery in the profile area), these changes were deemed by the research team and technical partners to not likely impact pilot trial outcomes in a meaningful way and given development cost, were prioritized for a later study.

App content was also reviewed by usability participants who provided specific recommendations for areas of the app that could be expanded (brain builders and daily discussion posts) and edited (collection stories and daily quests). The research team worked to address as many of these suggestions as possible before the pilot launch. Figures 1-6 provides screenshots of the fully developed *AllyQuest* app.

Figure 1. Profile page.

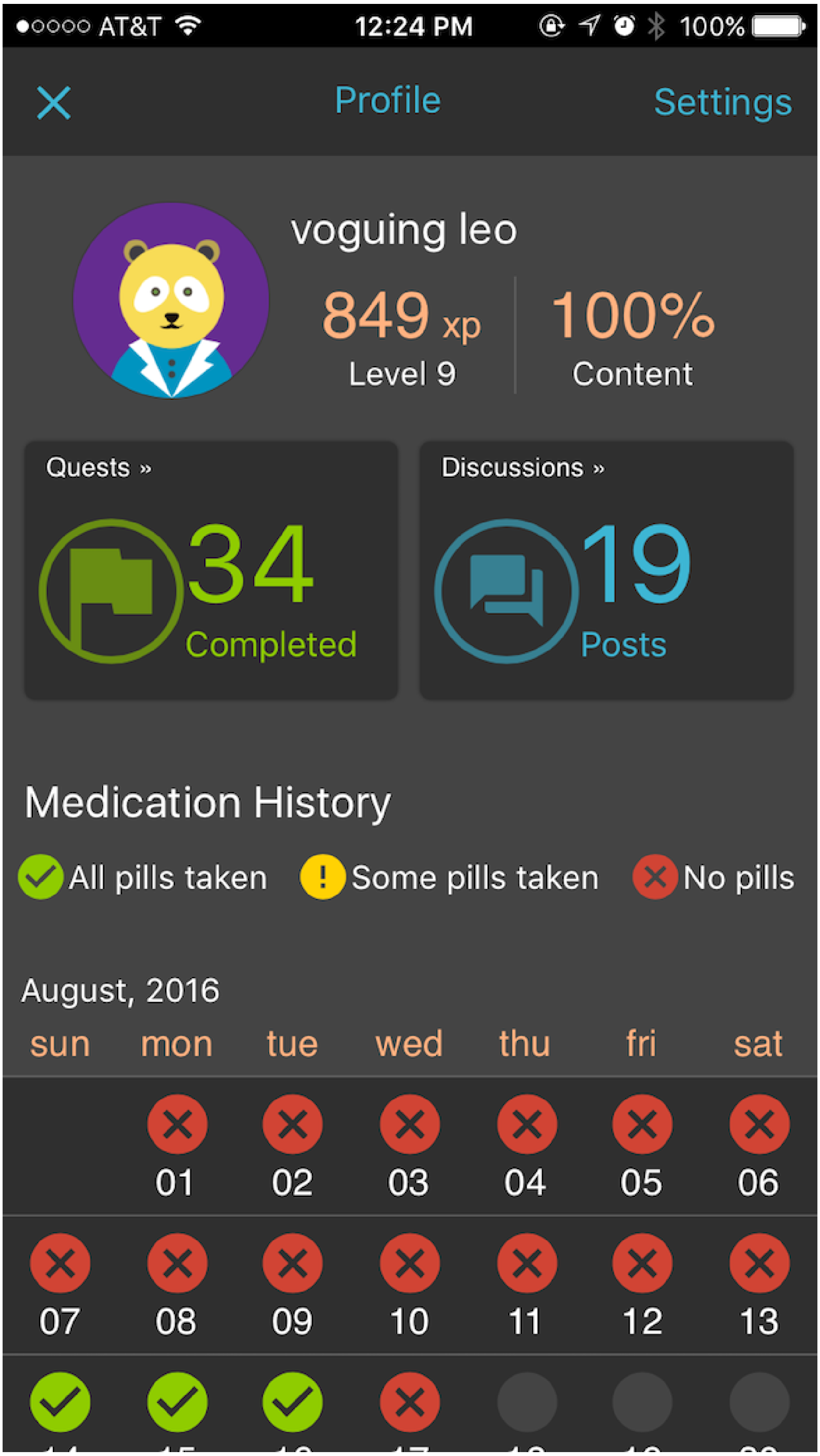
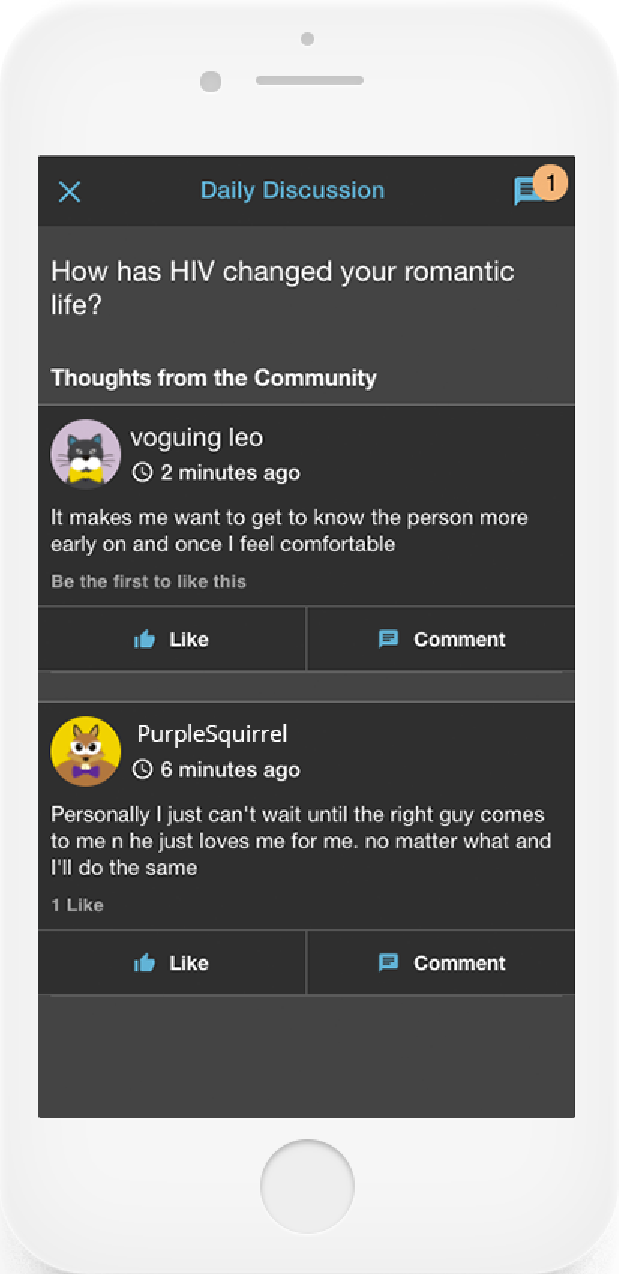


Figure 2. Daily discussion.



How has HIV changed your romantic life?

Thoughts from the Community

voguing leo

2 minutes ago

It makes me want to get to know the person more early on and once I feel comfortable

Be the first to like this

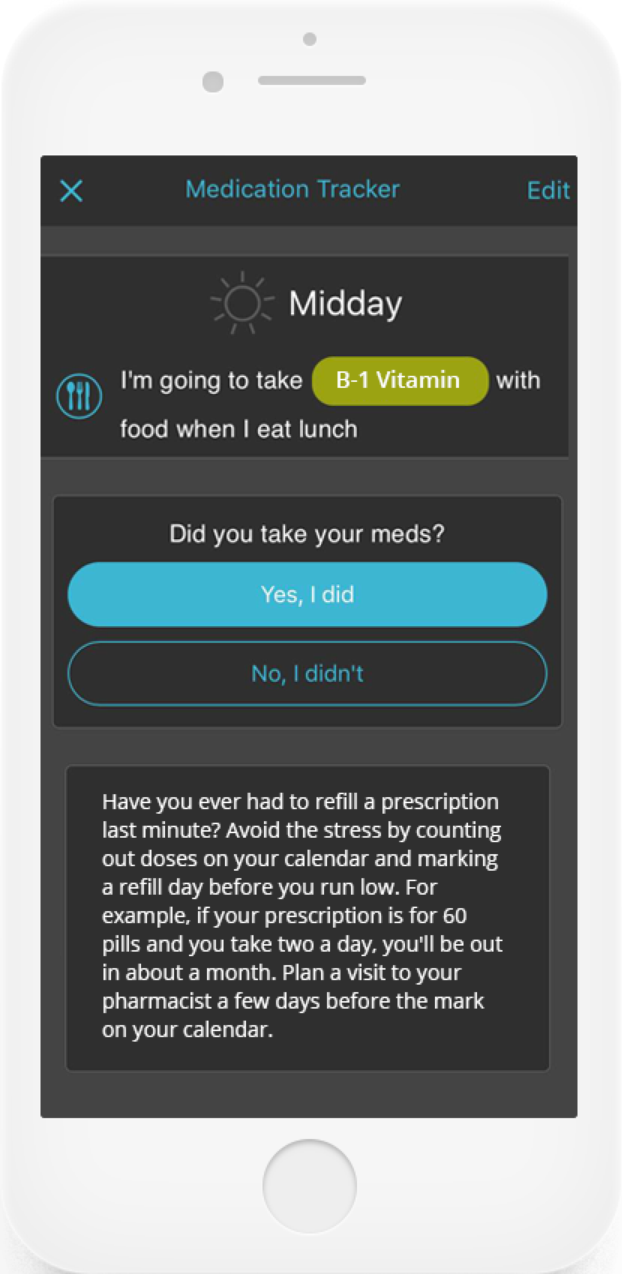
PurpleSquirrel

6 minutes ago

Personally I just can't wait until the right guy comes to me n he just loves me for me. no matter what and I'll do the same

1 Like

Figure 3. Medication tracker.

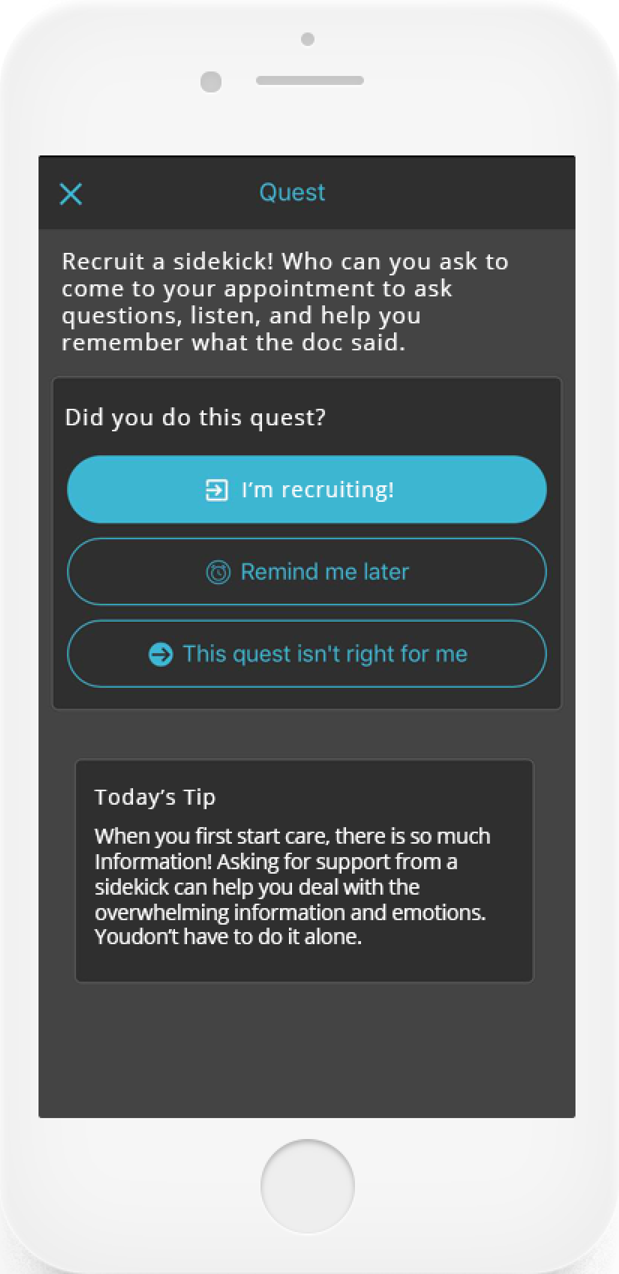


I'm going to take B-1 Vitamin with food when I eat lunch

Did you take your meds?

Have you ever had to refill a prescription last minute? Avoid the stress by counting out doses on your calendar and marking a refill day before you run low. For example, if your prescription is for 60 pills and you take two a day, you'll be out in about a month. Plan a visit to your pharmacist a few days before the mark on your calendar.

Figure 4. Daily quest.



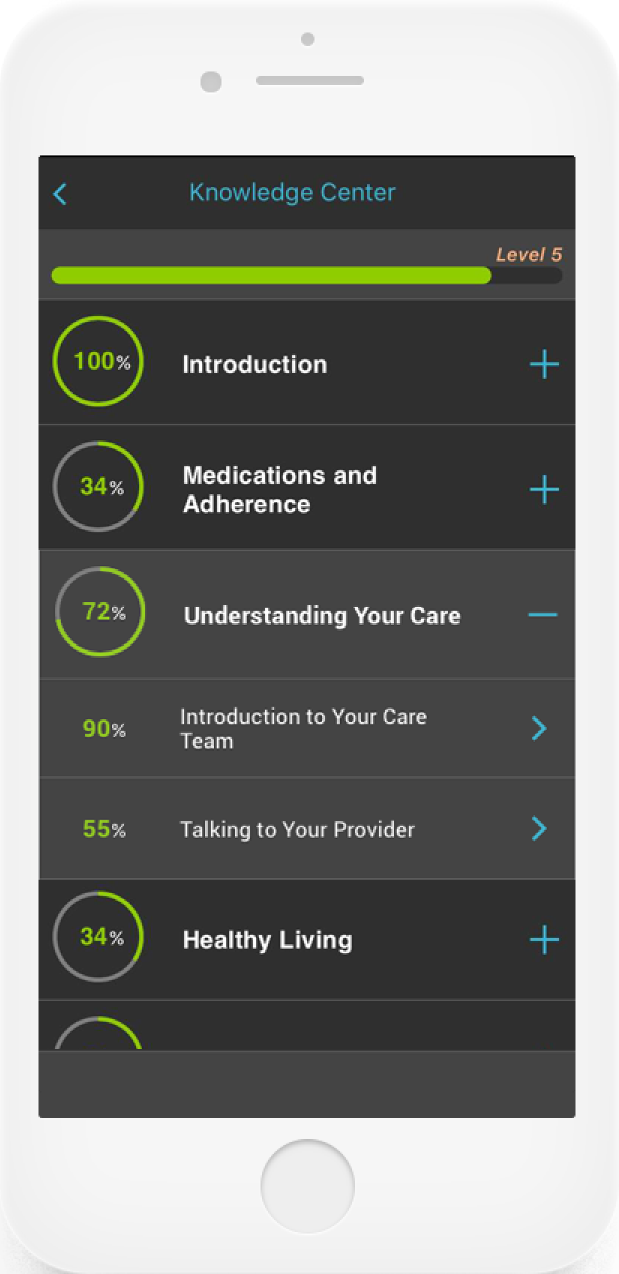
Recruit a sidekick! Who can you ask to come to your appointment to ask questions, listen, and help you remember what the doc said.

Did you do this quest?

Today's Tip

When you first start care, there is so much Information! Asking for support from a sidekick can help you deal with the overwhelming information and emotions. Youdon't have to do it alone.

Figure 5. Knowledge center article.



Knowledge Center

Level 5

100% Introduction

34% Medications and Adherence

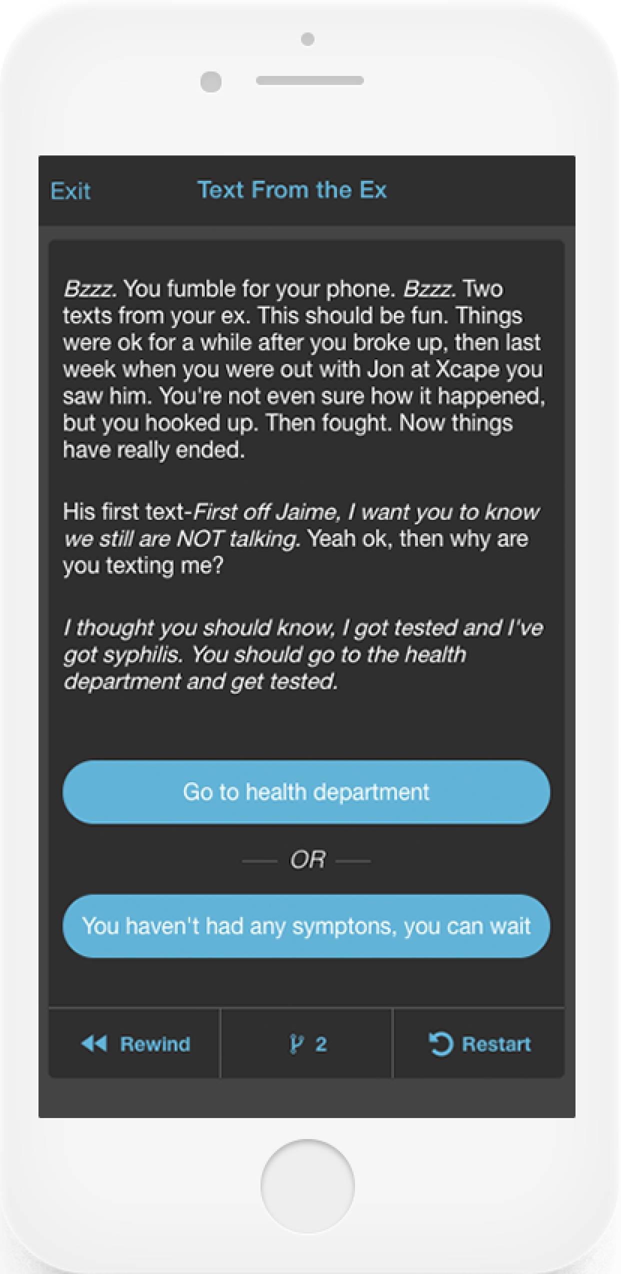
72% Understanding Your Care

90% Introduction to Your Care Team

55% Talking to Your Provider

34% Healthy Living

Figure 6. Narrative story.



Bzzz. You fumble for your phone. *Bzzz.* Two texts from your ex. This should be fun. Things were ok for a while after you broke up, then last week when you were out with Jon at Xcape you saw him. You're not even sure how it happened, but you hooked up. Then fought. Now things have really ended.

His first text-*First off Jaime, I want you to know we still are NOT talking.* Yeah ok, then why are you texting me?

I thought you should know, I got tested and I've got syphilis. You should go to the health department and get tested.

Table 2. Demographics *AllyQuest* pilot study participants, N=20.

Variable	Statistics
Age (years), mean (SD)	21.8 (1.55)
Race or ethnicity, n (%)	
Black	17 (85)
White	1 (5)
Hispanic or Latino	1 (5)
American Indian or Alaskan native	1 (5)
Education, n (%)	
Completed high school	12 (60)
Some college	3 (15)
Did not complete high school	5 (25)
Sexual identity, n (%)	
Gay	19 (95)
Bisexual	1 (5)
Employment, n (%)	
Currently employed	13 (65)
Homeless last 6 months, n (%)	
Yes	5 (25)
In HIV care, n (%)	
Yes	20 (100)
On HIV medication, n (%)	
Yes	19 (95)

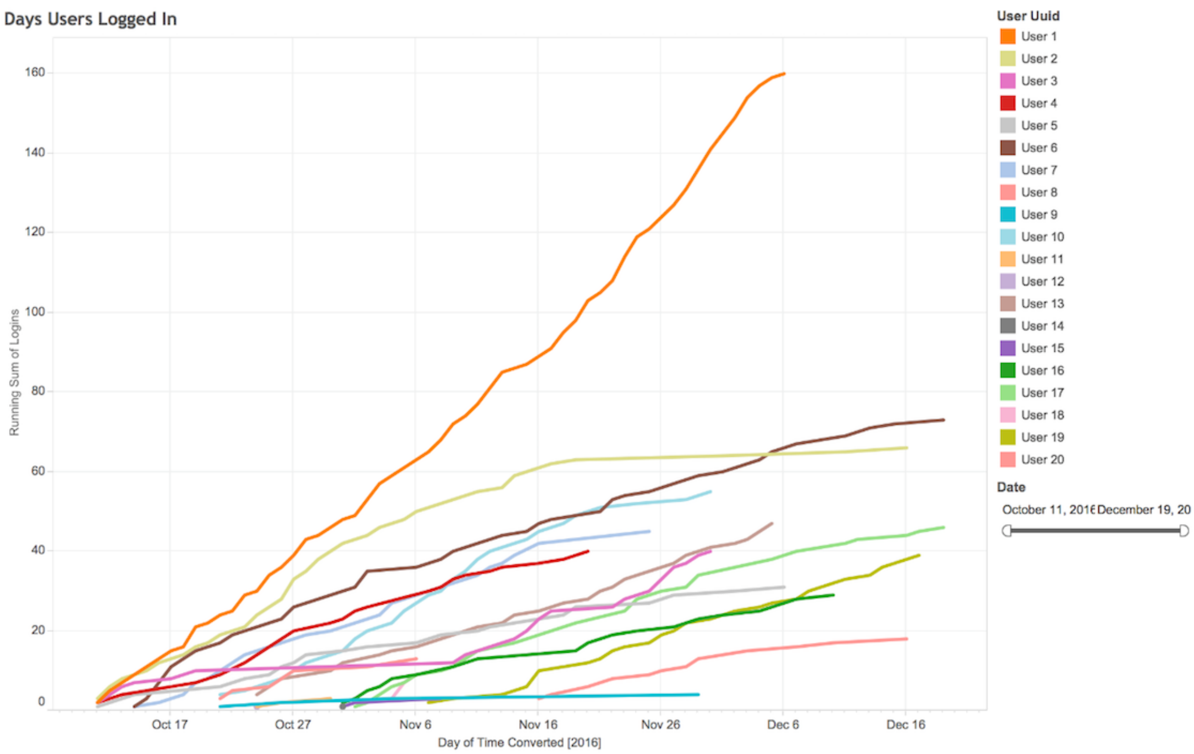
Pilot

The mean age was 21.8 years (range: 19-24), 95% (19/20) were nonwhite, 95% (19/20) identified as gay, 25% (5/20) had not completed high school, 65% (13/20) were currently employed, and 25% (5/20) reported homelessness in the past 6 months (Table 2). Most (16/20, 80%) participants had been diagnosed in the past year, all reported being engaged in care, and 95% (19/20) were currently prescribed ART. One-month retention was 85%, (17/20) though all pilot participants engaged in app use during the trial.

App Feasibility

The mean total time of app use was 158.4 min (SD 114.1), and range was 13 to 441 min. There was a mean of 21.2 days of use with a mean of 19.4 days of logging medication (Figure 7). App usage declined over the course of the trial, with a mean of 4.3, 3.4, 3.0, and 2.8 days of usage during weeks 1, 2, 3 and 4, respectively. Although participants were told that they only needed to use the app for 4 weeks, their access to *AllyQuest* was not discontinued until the final participant completed the trial. A total of 14 participants continued to use the app after their 4-week pilot trial period ended. There was a total of 17 knowledge center articles available during the pilot. Participants read a mean of 8.3 articles (range: 0-17). A total of 45 daily discussion questions were developed by the study team that appeared to participants on consecutive days during the pilot trial (eg, a participant enrolled on October 1 would see daily discussion topic #1, whereas a participant enrolling on October 7 would see would see daily discussion topic #7). If a participant did not log on, then they would not see the daily discussion topic that day but could navigate back to those conversations to comment. There were 222 posts to the daily discussion social wall, with a median of 5 posts (range: 1-11) for each daily discussion question. Most users (16/20) posted at least once during the 4-week pilot (median: 5.5 posts/person, range: 0-41 posts). The questions that received the most posts included “What is one goal you have for your health?” (11 posts), “How do you start the safe sex conversation?” (8 posts); “How do you deal with people who react badly to your status?” (7 posts), “How did you incorporate taking meds into your routine?” (7 posts), “What qualities do you appreciate in a healthcare provider?” (7 posts,) and “How has HIV changed your romantic life?” (7 posts).

Figure 7. Days and log-ins among participants in *AllyQuest* pilot trial.



App Acceptability

Acceptability ratings were high. Overall, participants found the app easy to use and navigate, not intrusive, and few reported technical issues (Table 3). The mean score on the CSQ-8 was 27.8 (SD 5.9). Most rated the quality of the app as excellent (n=10) or good (n=7), and overall, 15/17 were satisfied with the app. Overall, 16/17 participants felt they were getting the kind of service they wanted from the app, 14/17 felt that the app met most or almost all of their needs, 15/17 reported being mostly or very satisfied with the amount of help they received from the app, and 15/17 felt the app helped them deal more effectively with their HIV. Most (16/17) would recommend the app to a friend if they were in need of similar help, and 16/17 would use the app again.

Table 3. *AllyQuest* pilot study outcomes (n=17); 5-point Likert scale (1=strongly disagree-5=strongly agree).

Survey item	Mean (SD)	Median (Q1, Q3)	Minimum, maximum
Would use this app frequently	4.41 (1.18)	5.00 (4.00, 5.00)	1.00, 5.00
App was easy to use	4.76 (0.56)	5.00 (5.00, 5.00)	3.00, 5.00
Felt very confident using the app	4.47 (1.01)	5.00 (4.00, 5.00)	2.00, 5.00
App is accurate	4.53 (1.01)	5.00 (4.00, 5.00)	1.00, 5.00
App is dependable	4.53 (0.62)	5.00 (4.00, 5.00)	3.00, 5.00
Interaction with app is consistent	4.18 (0.95)	4.00 (4.00, 5.00)	2.00, 5.00
Found app unnecessarily complex	1.65 (0.86)	1.00 (1.00, 2.00)	1.00, 3.00
Would need technical assistance to use app	1.71 (1.16)	1.00 (1.00, 2.00)	1.00, 5.00
App features are well integrated	4.53 (0.80)	5.00 (4.00, 5.00)	2.00, 5.00
Most people could learn to use app quickly	4.65 (0.61)	5.00 (4.00, 5.00)	3.00, 5.00
Found app cumbersome to use	2.76 (1.64)	3.00 (1.00, 4.00)	1.00, 5.00
After versus before: connected to others with HIV	4.12 (0.70)	4.00 (4.00, 5.00)	3.00, 5.00
After versus before: ability to manage HIV	4.47 (0.72)	5.00 (4.00, 5.00)	3.00, 5.00
After versus before: ability to reliably take ART[a]	4.59 (0.71)	5.00 (4.00, 5.00)	3.00, 5.00
After versus before: knowledgeable about HIV	4.29 (0.77)	4.00 (4.00, 5.00)	3.00, 5.00

[a] ART: antiretroviral therapy.

Table 4. Correlations between app usage and HIV self-management outcomes, n=17. Outcome compared with before participating in the study.

App analytics	Connected to others with HIV, rho (*P* value)	Knowledgeable about condition, rho (*P* value)	Ability to manage condition, rho (*P* value)	Ability to reliably take medication, rho (*P* value)
Days logged in	.31 (.22)	.53 (.03)	.33 (.19)	.49 (.05)
Days logged medication	.34 (.20)	.42 (.10)	.19 (.48)	.41 (.11)
Total time on app	.13 (.61)	.42 (.09)	.04 (.86)	.15 (.58)
Articles read	.01 (.98)	.29 (.28)	−.10 (.72)	−.07 (.78)
Social wall posts	.34 (.18)	.48 (.52)	.40 (.12)	.35 (.17)
Daily quests	.32 (.22)	.43 (.88)	.12 (.63)	.36 (.16)

Qualitative exit interviews (n=17) identified areas for app improvement, including the need for additional tailoring and personalization. Overall, *AllyQuest* met the needs and expectations of its users, as expressed by some users:

> *But, you know, it did help me to overall accept that I have to take this prescription because it's only going to help me in the long run. So it was good as far as helping me remember to take the medicine.* [Participant 1004]
>
> *Being that I'm not much of a public speaker, in a sense, or an extroverted personality, the app kinda allowed me to, you know, bolster my genuine personal personality in a discrete manner, it was enjoyable.* [Participant 1008]
>
> *Next I know I was just on this thing all day. My friends are like, "You're always on your phone." I'm like, "Oh yeah, it's this app. I really can't talk about it's my personal life." They're like, "Oh, okay." So I'm just always on my phone.* [Participant 1007]

The daily discussion and medication tracker were users' favorite features. The discussions gave users a safe space to give and receive advice and made them feel less alone:

> *When I would read other people's comments on the little discussion panel thing—I would feel like I wasn't so alone...it made me feel like I was a part of a community that understood how hard it is dealing with something like that, because HIV is, like—it's something that's serious, and I always thought it was a death sentence because when I first found out I had it, I literally thought I was gonna die.* [Participant 1020]
>
> *You're talking to profiles and it's really, really cool cause I get a lot of feedback that I would need in my life because I don't talk about it with a lot of people. So this is a way that I get to keep it private and to myself and also get help. So it's really been amazing.* [Participant 1007]

The medication tracker helped users establish a medication strategy and normalize taking medications daily, as described by one user:

> *I've downloaded other apps where like a medicine tracker...it was just a very basic thing. And also I've had discussion forums on my phone where it just dove clean into something that was, like, I'm not gonna talk about that just yet. So I think the app was, like, right there in the middle. It offered the avenue to go deeper into the conversation and it offered just a safe space at the same time.* [Participant 1008]

Overall, users liked and trusted the health information on the app. Overall, they felt that the tone and content in the narratives was important and relevant. One participant described how these stories influenced his relationship with his partner:

> *Interviewer: How did the stories motivate you in the app?*
>
> *Participant: It kinda motivated me to be more open with my partner...I would keep a lotta things from them. Like, I would go through stressful times and I wouldn't tell them...but after seeing how a lotta that stuff played out, like in the [app] stories, I try to start opening up more, and I think we're gonna stay together for a long time.* [Participant 1002]

Participants did comment that they wanted more information regarding sexual health, relationships, mental health, and nutrition.

HIV Self-Management

Higher levels of app usage were positively correlated with HIV self-management outcomes, and there was a statistically significant ($P<.05$) positive association between the number of days logged into the app and knowledge and confidence in the ability to reliably take HIV medication (Table 4). Although statistical power was limited because of the small sample size, results are promising for a full intervention trial.

Discussion

Principal Findings

In this paper, we describe the development of a novel, theory-based ART adherence app for YMSM that showed strong acceptability, feasibility, and preliminary impact on HIV self-management outcomes. Prior work has established that technology-based interventions—particularly mHealth platforms—can provide tailored adherence interventions and allow YMSM to engage and connect with others [25-27]. Furthermore, daily mobile phone–based contact is acceptable to youth living with HIV and is associated with improved adherence [29]. Although several mHealth adherence studies for MSM are underway or under development [65,66], to our knowledge, there are no currently available interventions for

YMSM that include the range of features incorporated in *AllyQuest* or are poised to utilize technology in a similar highly innovative and engaging way.

HIV-positive YMSM were instrumental in informing all stages of *AllyQuest* development. Prior research with HIV-positive YMSM informed the inclusion of essential elements in the initial *AllyQuest* prototype, including information on both HIV-related issues and general health and wellness [64], anonymity and privacy features [58,59,64], medication reminders and tailored adherence strategies [59,63,67-69], and provision of social support and a connection with others [61,70]. Youth advisory board and usability participants provided further feedback that was incorporated in an agile way to allow for ongoing modifications and enhancements. The result of this process is a user-centered, highly engaging, multicomponent care support app.

AllyQuest accommodates different learning styles, motivations, and needs among YMSM through features including app-guided tailoring of content, personalized messages, and inclusion of game-based elements. Gamification uses game design components in nontraditional gaming contexts, thus providing opportunities for a greater level of engagement of participants in online behavioral interventions [60,71]. Interventions can utilize gamification to deliver highly engaging content and promote participant interactions both within and outside the intervention, thus increasing the potential for health behavior change. To optimize intervention engagement and impact, *AllyQuest* integrates health-related challenges, rewards, social connectivity, and “unlocking” character-driven narratives [60,71].

AllyQuest was built on an established platform developed by our technology partner, Ayogo. In a world of ever-shrinking resources, developing apps for ART adherence *de novo* may not take advantage of prior work done to improve treatment adherence in other disease states, thus failing to capitalize on lessons learned and software assets developed. Furthermore, collaborating with technology partners with established products allows for some degree of cost-sharing, ensuring that the app will be updated as needed and can be scaled up if proven effective. However, this process still requires active engagement with the target population to ensure adaptations are both developmentally and culturally congruent with their needs. Allowing sufficient time for iterative adjustments to the intervention is critical.

Understanding the full spectrum of app feasibility and acceptability before large-scale efficacy testing is essential. This entails measuring and subsequently aligning multiple streams of both in app (metrics of use and participant postings) and out of app (pre- and posttest surveys and qualitative exit interviews) data. Ensuring the app has a robust back-end data system to capture all app analytics is a crucial piece that should be prioritized early in development. In this study, feasibility and acceptability metrics aligned with our primary HIV self-management outcomes, increasing the likelihood that *AllyQuest* t could in fact impact long-term HIV ART adherence among HIV-positive YMSM. However, identifying additional strategies to ensure consistent and sustained app engagement should be considered. Integration of features that provide users who may not respond to technology-only solutions (eg, two-way text messaging or video counseling sessions with an adherence counselor) or stepped transition to in-person interventions should be considered in future studies.

Limitations

This study is not without limitations. Data on HIV self-management outcomes were self-reported and were only measured among participants who received the intervention. Due to the limited scope and length of this pilot study, enrolling a control group would not have been feasible, and measuring changes in biologic outcomes would not have been clinically meaningful. Statistical tests should be interpreted with caution, given the sample size. Findings may not be generalizable to YMSM from other sociodemographic backgrounds or geographic locations. Although we attempted to enroll diverse youth, the majority of participants were YMSM of color. Given the disproportionate impact of HIV among YMSM of color in the United States, these youth represent the population in highest need of interventions. Finally, we had 3 participants who did not complete their follow-up survey or interview, though all 3 participants logged on to the app and used it during the pilot study. Additional engagement and retention strategies will be important particularly when evaluating the impact of the app on out-of-care youth.

Conclusions

This small pilot trial confirmed that an interactive app is feasible and acceptable to YMSM as a tool to address ART adherence. Future work should build on the promising data from this trial to test *AllyQuest* in a larger, diverse sample to assess intervention efficacy for improving ART adherence and increasing sustained viral suppression. If a highly scalable technology such as *AllyQuest* could ultimately demonstrate effectiveness in implementation studies, it would be a powerful tool for realizing the individual and public health benefits of biomedical advances in prevention and care therapies. Furthermore, placing these tailored technologies in the hands of YMSM offers an approach to HIV self-management that may empower youth as they establish optimal HIV care engagement habits for the future.

Conflicts of Interest

Authors MS and EL were employees of Ayogo, Inc at the time of the study. There are no other conflicts to declare.

References

1. Centers for Disease Control and Prevention. 2016. HIV Surveillance Report, 2015 URL: https://www.cdc.gov/hiv/library/reports/hiv-surveillance.html [accessed 2018-02-26] [WebCite Cache ID 6xX30JSDz]
2. Centers for Disease Control and Prevention. 2015. HIV Surveillance Report: Diagnoses of HIV Infection in the United States and Dependent Areas, 2015 URL: https://www.cdc.gov/hiv/pdf/library/reports/surveillance/cdc-hiv-surveillance-report-2015-vol-27.pdf [accessed 2018-03-18] [WebCite Cache ID 6y0kyFQGY]
3. Wejnert C, Hess KL, Rose CE, Balaji A, Smith JC, Paz-Bailey G, NHBS Study Group. Age-Specific Race and Ethnicity Disparities in HIV Infection and Awareness Among Men Who Have Sex With Men--20 US Cities, 2008-2014. J Infect Dis 2016 Mar 01;213(5):776-783. [doi: 10.1093/infdis/jiv500] [Medline: 26486637]
4. Centers for Disease Control and Prevention. 2015. Monitoring selected national HIV prevention and care objectives by using HIV surveillance data - United States and 6 dependent areas - 2013 HIV Surveillance Supplemental Report 2015 Updated May URL: https://www.cdc.gov/hiv/pdf/library/reports/surveillance/cdc-hiv-surveillance-supplemental-report-vol-22-2.pdf [accessed 2018-03-18] [WebCite Cache ID 6y0l9KdHz]
5. Kapogiannis BX, Mayer K. The HIV continuum of care for adolescents and young adults (12-24 years) attending 13 urban US centers of the NICHD-ATN-CDC-HRSA SMILE collaborative. Presented at: IAS; 2015; Vancouver.
6. Kim SH, Gerver SM, Fidler S, Ward H. Adherence to antiretroviral therapy in adolescents living with HIV: systematic review and meta-analysis. AIDS 2014 Aug 24;28(13):1945-1956 [FREE Full text] [doi: 10.1097/QAD.0000000000000316] [Medline: 24845154]
7. Wilson PA, Kahana SY, Fernandez MI, Harper GW, Mayer K, Wilson CM, et al. Sexual risk behavior among virologically detectable human immunodeficiency virus-infected young men who have sex with men. JAMA Pediatr 2016 Feb;170(2):125-131 [FREE Full text] [doi: 10.1001/jamapediatrics.2015.3333] [Medline: 26641367]
8. MacDonell K, Naar-King S, Huszti H, Belzer M. Barriers to medication adherence in behaviorally and perinatally infected youth living with HIV. AIDS Behav 2013 Jan;17(1):86-93 [FREE Full text] [doi: 10.1007/s10461-012-0364-1] [Medline: 23142855]
9. Houston E, Fominaya AW. Antiretroviral therapy adherence in a sample of men with low socioeconomic status: the role of task-specific treatment self-efficacy. Psychol Health Med 2015;20(8):896-905. [doi: 10.1080/13548506.2014.986137] [Medline: 25439192]
10. Kuhns LM, Hotton AL, Garofalo R, Muldoon AL, Jaffe K, Bouris A, et al. An index of multiple psychosocial, syndemic conditions is associated with antiretroviral medication adherence among HIV-positive youth. AIDS Patient Care STDS 2016 Apr;30(4):185-192 [FREE Full text] [doi: 10.1089/apc.2015.0328] [Medline: 27028184]
11. Murphy DA, Marelich WD, Hoffman D, Steers WN. Predictors of antiretroviral adherence. AIDS Care 2004 May;16(4):471-484. [doi: 10.1080/09540120410001683402] [Medline: 15203415]
12. Shacham E, Estlund AL, Tanner AE, Presti R. Challenges to HIV management among youth engaged in HIV care. AIDS Care 2017 Feb;29(2):189-196. [doi: 10.1080/09540121.2016.1204422] [Medline: 27397139]
13. White JM, Gordon JR, Mimiaga MJ. The role of substance use and mental health problems in medication adherence among HIV-infected MSM. LGBT Health 2014 Dec;1(4):319-322. [doi: 10.1089/lgbt.2014.0020] [Medline: 26789859]
14. Gross IM, Hosek S, Richards MH, Fernandez MI. Predictors and profiles of antiretroviral therapy adherence among African American adolescents and young adult males living with HIV. AIDS Patient Care STDS 2016 Jul;30(7):324-338 [FREE Full text] [doi: 10.1089/apc.2015.0351] [Medline: 27410496]
15. Rudy BJ, Murphy DA, Harris DR, Muenz L, Ellen J, Adolescent Trials Network for HIV/AIDS Interventions. Patient-related risks for nonadherence to antiretroviral therapy among HIV-infected youth in the United States: a study of prevalence and interactions. AIDS Patient Care STDS 2009 Mar;23(3):185-194 [FREE Full text] [doi: 10.1089/apc.2008.0162] [Medline: 19866536]
16. Macdonell KE, Naar-King S, Murphy DA, Parsons JT, Harper GW. Predictors of medication adherence in high risk youth of color living with HIV. J Pediatr Psychol 2010 Jul;35(6):593-601 [FREE Full text] [doi: 10.1093/jpepsy/jsp080] [Medline: 19755495]
17. Bogart LM, Landrine H, Galvan FH, Wagner GJ, Klein DJ. Perceived discrimination and physical health among HIV-positive Black and Latino men who have sex with men. AIDS Behav 2013 May;17(4):1431-1441 [FREE Full text] [doi: 10.1007/s10461-012-0397-5] [Medline: 23297084]
18. Friedman MR, Stall R, Silvestre AJ, Wei C, Shoptaw S, Herrick A, et al. Effects of syndemics on HIV viral load and medication adherence in the multicentre AIDS cohort study. AIDS 2015 Jun 01;29(9):1087-1096 [FREE Full text] [doi: 10.1097/QAD.0000000000000657] [Medline: 25870981]
19. Risher KA, Kapoor S, Daramola AM, Paz-Bailey G, Skarbinski J, Doyle K, et al. Challenges in the evaluation of interventions to improve engagement along the HIV care continuum in the United States: a systematic review. AIDS Behav 2017 Jul;21(7):2101-2123. [doi: 10.1007/s10461-017-1687-8] [Medline: 28120257]
20. Chaiyachati KH, Ogbuoji O, Price M, Suthar AB, Negussie EK, Bärnighausen T. Interventions to improve adherence to antiretroviral therapy: a rapid systematic review. AIDS 2014 Mar;28 Suppl 2:S187-S204. [doi: 10.1097/QAD.0000000000000252] [Medline: 24849479]
21. Charania MR, Marshall KJ, Lyles CM, Crepaz N, Kay LS, Koenig LJ, HIV/AIDS Prevention Research Synthesis (PRS) Team. Identification of evidence-based interventions for promoting HIV medication adherence: findings from a systematic

review of U.S.-based studies, 1996-2011. AIDS Behav 2014 Apr;18(4):646-660 [FREE Full text] [doi: 10.1007/s10461-013-0594-x] [Medline: 24043269]

22. Noar SM, Benac CN, Harris MS. Does tailoring matter? Meta-analytic review of tailored print health behavior change interventions. Psychol Bull 2007 Jul;133(4):673-693. [doi: 10.1037/0033-2909.133.4.673] [Medline: 17592961]
23. Kreuter MW, Wray RJ. Tailored and targeted health communication: strategies for enhancing information relevance. Am J Health Behav 2003 Dec;27 Suppl 3:S227-S232. [Medline: 14672383]
24. Krebs P, Prochaska JO, Rossi JS. A meta-analysis of computer-tailored interventions for health behavior change. Prev Med 2010;51(3-4):214-221 [FREE Full text] [doi: 10.1016/j.ypmed.2010.06.004] [Medline: 20558196]
25. Hightow-Weidman LB, Muessig KE, Bauermeister J, Zhang C, LeGrand S. Youth, technology, and HIV: recent advances and future directions. Curr HIV/AIDS Rep 2015 Dec;12(4):500-515. [doi: 10.1007/s11904-015-0280-x] [Medline: 26385582]
26. LeGrand S, Muessig KE, McNulty T, Soni K, Knudtson K, Lemann A, et al. Epic Allies: development of a gaming app to improve antiretroviral therapy adherence among young HIV-positive men who have sex with men. JMIR Serious Games 2016 May 13;4(1):e6 [FREE Full text] [doi: 10.2196/games.5687] [Medline: 27178752]
27. Muessig KE, Nekkanti M, Bauermeister J, Bull S, Hightow-Weidman LB. A systematic review of recent smartphone, Internet and Web 2.0 interventions to address the HIV continuum of care. Curr HIV/AIDS Rep 2015 Jan 28;12(1):173-190. [doi: 10.1007/s11904-014-0239-3] [Medline: 25626718]
28. Shaw RJ, Steinberg DM, Zullig LL, Bosworth HB, Johnson CM, Davis LL. mHealth interventions for weight loss: a guide for achieving treatment fidelity. J Am Med Inform Assoc 2014;21(6):959-963. [doi: 10.1136/amiajnl-2013-002610] [Medline: 24853065]
29. Belzer ME, Kolmodin MacDonell K, Clark LF, Huang J, Olson J, Kahana SY, Adolescent Medicine Trials Network for HIV/AIDS Interventions. Acceptability and feasibility of a cell phone support intervention for youth living with HIV with nonadherence to antiretroviral therapy. AIDS Patient Care STDS 2015 Jun;29(6):338-345 [FREE Full text] [doi: 10.1089/apc.2014.0282] [Medline: 25928772]
30. Garofalo R, Kuhns LM, Hotton A, Johnson A, Muldoon A, Rice D. A randomized controlled trial of personalized text message reminders to promote medication adherence among HIV-positive adolescents and young adults. AIDS Behav 2016 May;20(5):1049-1059. [doi: 10.1007/s10461-015-1192-x] [Medline: 26362167]
31. Johnson MO, Charlebois E, Morin SF, Remien RH, Chesney MA, National Institute of Mental Health Healthy Living Project Team. Effects of a behavioral intervention on antiretroviral medication adherence among people living with HIV: the healthy living project randomized controlled study. J Acquir Immune Defic Syndr 2007 Dec 15;46(5):574-580 [FREE Full text] [Medline: 18193499]
32. Wilton L, Herbst JH, Coury-Doniger P, Painter TM, English G, Alvarez ME, et al. Efficacy of an HIV/STI prevention intervention for black men who have sex with men: findings from the Many Men, Many Voices (3MV) project. AIDS Behav 2009 Jun;13(3):532-544. [doi: 10.1007/s10461-009-9529-y] [Medline: 19267264]
33. Bandura A. Health promotion by social cognitive means. Health Educ Behav 2004 Apr;31(2):143-164. [doi: 10.1177/1090198104263660] [Medline: 15090118]
34. Bandura A. Human agency in social cognitive theory. Am Psychol 1989 Sep;44(9):1175-1184. [Medline: 2782727]
35. Fix GM, Houston TK, Barker AM, Wexler L, Cook N, Volkman JE, et al. A novel process for integrating patient stories into patient education interventions: incorporating lessons from theater arts. Patient Educ Couns 2012 Sep;88(3):455-459. [doi: 10.1016/j.pec.2012.06.012] [Medline: 22770814]
36. Houston TK, Cherrington A, Coley HL, Robinson KM, Trobaugh JA, Williams JH, et al. The art and science of patient storytelling-harnessing narrative communication for behavioral interventions: the ACCE project. J Health Commun 2011 Aug;16(7):686-697. [doi: 10.1080/10810730.2011.551997] [Medline: 21541875]
37. Petraglia J. Narrative intervention in behavior and public health. J Health Commun 2007;12(5):493-505. [doi: 10.1080/10810730701441371] [Medline: 17710598]
38. Petraglia J, Galavotti C, Harford N, Pappas-DeLuca KA, Mooki M. Applying behavioral science to behavior change communication: the pathways to change tools. Health Promot Pract 2007 Oct;8(4):384-393. [doi: 10.1177/1524839907301402] [Medline: 17804825]
39. Fogg BJ. Persuasive technology: using computers to change what we think and do. Ubiquity 2002 Dec 01;2002(December):2. [doi: 10.1145/764008.763957]
40. Abraham C, Sheeran P, Orbell S. Can social cognitive models contribute to the effectiveness of HIV-preventive behavioural interventions? A brief review of the literature and a reply to Joffe (1996; 1997) and Fife-Schaw (1997). Br J Med Psychol 1998 Sep;71 (Pt 3):297-310. [Medline: 9733424]
41. Gucciardi E, Jean-Pierre N, Karam G, Sidani S. Designing and delivering facilitated storytelling interventions for chronic disease self-management: a scoping review. BMC Health Serv Res 2016 Jul 11;16:249 [FREE Full text] [doi: 10.1186/s12913-016-1474-7] [Medline: 27401836]
42. Houston TK, Allison JJ, Sussman M, Horn W, Holt CL, Trobaugh J, et al. Culturally appropriate storytelling to improve blood pressure: a randomized trial. Ann Intern Med 2011 Jan 18;154(2):77-84. [doi: 10.7326/0003-4819-154-2-201101180-00004] [Medline: 21242364]

43. McQueen A, Kreuter MW, Kalesan B, Alcaraz KI. Understanding narrative effects: the impact of breast cancer survivor stories on message processing, attitudes, and beliefs among African American women. Health Psychol 2011 Nov;30(6):674-682 [FREE Full text] [doi: 10.1037/a0025395] [Medline: 21895370]
44. Hightow-Weidman LB, Muessig K, Srivatsa M, Lawrence E, LeGrand S, Kirschke-Schwartz H, et al. AllyQuest: Engaging HIV Young MSM in Care and Improving Adherence through a Social Networking and Gamified Smartphone App. Presented at: 9th International AIDS Society Conference on HIV Science (IAS); 2017; Paris, France.
45. Hightow-Weidman LB, Muessig KE, Pike EC, LeGrand S, Baltierra N, Rucker AJ, et al. HealthMpowerment.org: building community through a mobile-optimized, online health promotion intervention. Health Educ Behav 2015 Aug;42(4):493-499. [doi: 10.1177/1090198114562043] [Medline: 25588932]
46. Usability. URL: https://www.usability.gov/ [accessed 2018-02-26] [WebCite Cache ID 6xX3v6pYK]
47. Rubin J, Chisnell D. Handbook of Usability Testing: How to Plan, Design, and Conduct Effective Tests Second Edition. Indianapolis, IN: Wiley Publishing, Inc; 2008.
48. Olmsted-Hawala EL, Murphy ED, Hawala S, Ashenfelter KT. Does age make a difference? Presented at: Proceedings from the Society for Technical Communication Summit; 2012; Chicago, IL.
49. Richardson S, Mishuris R, O'Connell A, Feldstein D, Hess R, Smith P, et al. “Think aloud” and “Near live” usability testing of two complex clinical decision support tools. Int J Med Inform 2017 Oct;106:1-8. [doi: 10.1016/j.ijmedinf.2017.06.003] [Medline: 28870378]
50. Brooke J. A quick and dirty usability scale. In: Weerdmester B, McClelland I, editors. Usability Evaluation in Industry. London: Taylor and Francis; 1996:189-194.
51. Matsubara C, Green J, Astorga LT, Daya EL, Jervoso HC, Gonzaga EM, et al. Reliability tests and validation tests of the client satisfaction questionnaire (CSQ-8) as an index of satisfaction with childbirth-related care among Filipino women. BMC Pregnancy Childbirth 2013;13:235 [FREE Full text] [doi: 10.1186/1471-2393-13-235] [Medline: 24341288]
52. Cordova D, Alers-Rojas F, Lua FM, Bauermeister J, Nurenberg R, Ovadje L, et al. The usability and acceptability of an Aaolescent mHealth HIV/STI and drug abuse preventive intervention in primary care. Behav Med 2016 May 25:1-12. [doi: 10.1080/08964289.2016.1189396] [Medline: 27223646]
53. Tsai LLY, McNamara RJ, Dennis SM, Moddel C, Alison JA, McKenzie DK, et al. Satisfaction and experience with a supervised home-based real-time videoconferencing telerehabilitation exercise program in people with chronic obstructive pulmonary disease (COPD). Int J Telerehabil 2016;8(2):27-38 [FREE Full text] [doi: 10.5195/ijt.2016.6213] [Medline: 28775799]
54. Nobis S, Lehr D, Ebert DD, Baumeister H, Snoek F, Riper H, et al. Efficacy of a web-based intervention with mobile phone support in treating depressive symptoms in adults with type 1 and type 2 diabetes: a randomized controlled trial. Diabetes Care 2015 May;38(5):776-783. [doi: 10.2337/dc14-1728] [Medline: 25710923]
55. Donker T, Bennett K, Bennett A, Mackinnon A, van Straten A, Cuijpers P, et al. Internet-delivered interpersonal psychotherapy versus internet-delivered cognitive behavioral therapy for adults with depressive symptoms: randomized controlled noninferiority trial. J Med Internet Res 2013;15(5):e82 [FREE Full text] [doi: 10.2196/jmir.2307] [Medline: 23669884]
56. Astivia OL, Zumbo BD. Population models and simulation methods: the case of the Spearman rank correlation. Br J Math Stat Psychol 2017 Jan 31;70(3):347-367. [doi: 10.1111/bmsp.12085] [Medline: 28140458]
57. Janosky JE. Pearson correlation coefficients vs reliability coefficients. J Am Diet Assoc 1991 Aug;91(8):912-913. [Medline: 1894897]
58. Muessig KE, Pike EC, Fowler B, LeGrand S, Parsons JT, Bull SS, et al. Putting prevention in their pockets: developing mobile phone-based HIV interventions for black men who have sex with men. AIDS Patient Care STDS 2013 Apr;27(4):211-222 [FREE Full text] [doi: 10.1089/apc.2012.0404] [Medline: 23565925]
59. Senn TE, Braksmajer A, Coury-Doniger P, Urban MA, Carey MP. Mobile technology use and desired technology-based intervention characteristics among HIV+ black men who have sex with men. AIDS Care 2017 Apr;29(4):423-427. [doi: 10.1080/09540121.2016.1220479] [Medline: 27535069]
60. Cugelman B. Gamification: what it is and why it matters to digital health behavior change developers. JMIR Serious Games 2013;1(1):e3 [FREE Full text] [doi: 10.2196/games.3139] [Medline: 25658754]
61. Senn TE, Braksmajer A, Coury-Doniger P, Urban MA, Rossi A, Carey MP. Development and preliminary pilot testing of a peer support text messaging intervention for HIV-infected black men who have sex with men. J Acquir Immune Defic Syndr 2017 Feb 01;74 Suppl 2:S121-S127. [doi: 10.1097/QAI.0000000000001241] [Medline: 28079722]
62. Saberi P, Johnson MO. Technology-based self-care methods of improving antiretroviral adherence: a systematic review. PLoS One 2011;6(11):e27533. [doi: 10.1371/journal.pone.0027533] [Medline: 22140446]
63. Lewis MA, Uhrig JD, Bann CM, Harris JL, Furberg RD, Coomes C, et al. Tailored text messaging intervention for HIV adherence: a proof-of-concept study. Health Psychol 2013 Mar;32(3):248-253. [doi: 10.1037/a0028109] [Medline: 22545972]
64. Saberi P, Siedle-Khan R, Sheon N, Lightfoot M. The use of mobile health applications among youth and young adults living with HIV: focus group findings. AIDS Patient Care STDS 2016 Jun;30(6):254-260 [FREE Full text] [doi: 10.1089/apc.2016.0044] [Medline: 27214751]

65. Tanner AE, Mann L, Song E, Alonzo J, Schafer K, Arellano E, et al. weCARE: a social media-based intervention designed to increase HIV care linkage, retention, and health outcomes for racially and ethnically diverse young MSM. AIDS Educ Prev 2016 Jun;28(3):216-230 [FREE Full text] [doi: 10.1521/aeap.2016.28.3.216] [Medline: 27244190]
66. Muessig KE, LeGrand S, Horvath KJ, Bauermeister JA, Hightow-Weidman LB. Recent mobile health interventions to support medication adherence among HIV-positive MSM. Curr Opin HIV AIDS 2017 Sep;12(5):432-441. [doi: 10.1097/COH.0000000000000401] [Medline: 28639990]
67. Dowshen N, Kuhns LM, Johnson A, Holoyda BJ, Garofalo R. Improving adherence to antiretroviral therapy for youth living with HIV/AIDS: a pilot study using personalized, interactive, daily text message reminders. J Med Internet Res 2012;14(2):e51 [FREE Full text] [doi: 10.2196/jmir.2015] [Medline: 22481246]
68. Schnall R, Bakken S, Rojas M, Travers J, Carballo-Dieguez A. mHealth technology as a persuasive tool for treatment, care and management of persons living with HIV. AIDS Behav 2015 Jun;19 Suppl 2:81-89. [doi: 10.1007/s10461-014-0984-8] [Medline: 25572830]
69. Schnall R, Mosley JP, Iribarren SJ, Bakken S, Carballo-Diéguez A, Brown IW. Comparison of a user-centered design, self-management app to existing mHealth apps for persons living with HIV. JMIR Mhealth Uhealth 2015;3(3):e91 [FREE Full text] [doi: 10.2196/mhealth.4882] [Medline: 26385783]
70. Horvath KJ, Alemu D, Danh T, Baker JV, Carrico AW. Creating effective mobile phone apps to optimize antiretroviral therapy adherence: perspectives from stimulant-using HIV-positive men who have sex with men. JMIR Mhealth Uhealth 2016 Apr 15;4(2):e48 [FREE Full text] [doi: 10.2196/mhealth.5287] [Medline: 27084049]
71. King D, Greaves F, Exeter C, Darzi A. 'Gamification': influencing health behaviours with games. J R Soc Med 2013 Mar;106(3):76-78. [doi: 10.1177/0141076813480996] [Medline: 23481424]

Abbreviations

ART: antiretroviral therapy
CSQ-8: Client Satisfaction Questionnaire-8
FBM: Fogg Behavior Model
mHealth: mobile health
MSM: men who have sex with men
SCT: social cognitive theory
SUS: System Usability Scale
YMSM: young men who have sex with men

Edited by G Eysenbach; submitted 08.09.17; peer-reviewed by R Schnall, K Card; comments to author 28.09.17; revised version received 31.10.17; accepted 07.12.17; published 30.04.18.

Please cite as:
Hightow-Weidman L, Muessig K, Knudtson K, Srivatsa M, Lawrence E, LeGrand S, Hotten A, Hosek S
A Gamified Smartphone App to Support Engagement in Care and Medication Adherence for HIV-Positive Young Men Who Have Sex With Men (AllyQuest): Development and Pilot Study
JMIR Public Health Surveill 2018;4(2):e34
URL: http://publichealth.jmir.org/2018/2/e34/
doi:10.2196/publichealth.8923
PMID:29712626

Original Paper

Recommendations for the Development of a Mobile HIV Prevention Intervention for Men Who Have Sex With Men and Hijras in Mumbai: Qualitative Study

Shruta Rawat[1*], MA; J Michael Wilkerson[2*], MPH, PhD; Sylvia M Lawler[2*], MA; Pallav Patankar[1*], MBA; BR Simon Rosser[3*], MPH, PhD; Kanjani Shukla[4], MPH; Seyram Butame[2], MPH, PhD; Maria L Ekstrand[5*], PhD

[1]The Humsafar Trust, Mumbai, India
[2]School of Public Health, University of Texas Health Science Center at Houston, Houston, TX, United States
[3]School of Public Health, University of Minnesota, Minneapolis, MN, United States
[4]Minnesota Department of Health, Minneapolis, MN, United States
[5]University of California San Francisco Center for AIDS Prevention Studies, San Francisco, CA, United States
[*]these authors contributed equally

Corresponding Author:
J Michael Wilkerson, MPH, PhD
School of Public Health
University of Texas Health Science Center at Houston
7000 Fannin St, Ste 2620
Houston, TX,
United States
Phone: 1 713 500 9974
Email: Johnny.M.Wilkerson@uth.tmc.edu

Abstract

Background: As Internet and mobile phone use expands in India, there is an opportunity to develop mobile health (mHealth) interventions for marginalized populations, including men who have sex with men (MSM) and hijras (transgender women), hesitant to access traditional health care systems.

Objective: The purpose of this study was to determine if an mHealth intervention was acceptable to MSM and hijras living in Mumbai, and if so, what features would be useful in targeting the prevention of HIV acquisition and to increase the quality of life among persons living with HIV/AIDS.

Methods: Data from 4 focus groups with MSM and interviews with 4 hijras, 10 health service providers, and 8 mHealth developers were thematically analyzed.

Results: Once the need for an mHealth intervention was confirmed, comments about features were organized into 3 themes: content, interface, and retention. Content subthemes included providing sex education for younger community members, providing information about STIs, and providing information and social support for persons living with HIV. Interface subthemes included presenting content using pictures; using videos to present stories of role models; using push notifications for testing, appointment, and medication reminders; using geolocation to link to just-in-time services; and using telemedicine to increase access to health service providers and community services. The 5 retention subthemes included keeping it fun, using gaming mechanics, developing content in regional languages, protecting confidentiality, and linking to social networking apps.

Conclusions: These findings may help inform mHealth development in India.

(JMIR Public Health Surveill 2018;4(2):e46) doi:10.2196/publichealth.9088

KEYWORDS

health promotion; health seeking; gays and lesbians; Internet; HIV/AIDS

Introduction

The National AIDS Control Organization estimates the prevalence of HIV among adults in India as 0.3% [1]. Among men who have sex with men (MSM) and hijras (transgender women), the prevalence of HIV is estimated to be 4.3% and 7.5%, respectively [1,2]. These relatively high rates of HIV among MSM and hijras have been attributed to individual-level risk behaviors like high rates of sexual concurrency (wherein individuals have multiple overlapping sexual partners) and low rates of condom use [3,4]. Additionally, MSM and hijras often face stigma, discrimination, and, in some cases, criminalization [5-7]. Due to fear of social recrimination, MSM and hijras may be unwilling to engage with health service providers, leaving them with unmet health care needs [7]. Some evidence suggests interventions using mobile health (mHealth) technology may increase access to health care in developing countries such as India [8]. Therefore, Indian MSM and hijras may also benefit from mHealth interventions.

In developing countries, including India, mHealth interventions have historically made use of short message service (SMS) technology [9-14]. However, the mobile technology market in India is complicated by governmental regulations, including a National Customer Preference Register, that limit the number of SMS messages sent and received per day [15]. Due to these limitations, researchers investigating the effects of SMS mHealth interventions in India have obtained mixed results [16,17]. It is estimated that in 2018 there will be 530 million mobile phone users in India, nearly twice that of the United States [18]. Given the challenges of harnessing SMS messaging for health interventions, researchers instead should consider the burgeoning mobile phone market within India as a means of advancing the field of mHealth, particularly in support of the unique health needs of MSM and hijras [19]. The difficulties in harnessing SMS technology for mHealth interventions make mobile phone technology such as Internet-enabled apps a viable option for public health interventionists.

In Western countries, many mHealth interventions capitalize on mobile phones to deliver HIV prevention messaging and needs assessments for high-risk MSM. Preferred content often includes information about HIV and sexually transmitted infections (STIs), proper condom use, behavioral risk reduction strategies, and partner communication strategies [20,21]. For persons living with HIV/AIDS (PLHA), preferred content includes information about health care providers and treatment, especially regarding linkage and retention into care and medication adherence [22,23]. Additional features include use of geolocation of HIV testing centers; automated reminders for testing, appointment, and medication adherence; use of existing public or private social networking apps and chat rooms to increase social support; links to role model videos; self-assessment tools; access to sexual health experts; referrals to health service providers; and telemedicine [22,24-27].

The successes and innovations of mobile phone–based mHealth interventions in Western populations make a strong case for developing such interventions for MSM and hijras in India. Thus, in this paper we report qualitative findings from a formative research study with high-risk Indian MSM and hijras, health service providers with MSM and hijra clientele, and mHealth developers to determine what content and features they would like to be included in an mHealth intervention to prevent new HIV infections and increase the quality of life among PLHA. To the best of our knowledge, there is no information about the preferred features of an mHealth intervention targeting high-risk MSM and hijras in India.

Methods

Study Population and Recruitment

We used a convenience sample to recruit study participants based on key informants identified by the research team and The Humsafar Trust, one of India’s largest nongovernmental organizations that advocates for the rights of sexual and gender minority people. We used snowball sampling [22,25] to recruit MSM into focus groups. Approximately 6 individuals were in each focus group. Hijras and health service providers were asked to participate in individual interviews. Eligibility criteria for MSM and hijras included being 18 years or older, living in Mumbai, and having used Internet-enabled technology (ie, laptop, desktop, tablet, or mobile phone) to meet male sex partners [28]. Health service providers included medical doctors, nurses, mental health providers, and outreach workers who had a significant number of MSM and hijra clients and who felt confident in their ability to speak about using technology to interact with patients. mHealth developers were recruited from an mHealth conference listserv. Conference organizers allowed the research team access to the email addresses of all attendees. Developers who had experience working with Indian populations were asked to contact the research team. Upon completion of focus group or interview, participants received Rs 300 (US $7). All study procedures were approved by the institutional review boards of all authors’ home institutions and the Indian Government.

The median age of the 24 MSM who participated in focus groups was 27 (interquartile range, IQR, 24 to 33) years. Most of them (21/24, 88%) were college educated. Nearly all the MSM (23/24, 99%) identified as gay or homosexual. About a third of them (9/24, 38%) were out about their sexual or gender identity to most or all the people they knew, 21% (5/24) were out to about half the people they knew, and 42% (10/24) of MSM participants were out to only a few people or no one at all.

The median age of the 4 hijras who participated in individual interviews was 25 (IQR 22 to 29) years. Three were college educated and all identified as hijras. Three were out to most of the people they knew as a hijra, while 1 was out to about half the people she knew.

The median age of the 10 health service providers who participated in individual interviews was 31 (IQR 28 to 38) years. Half of them (5/10, 50%) were college educated; 8 of the health service providers identified as men, 1 as a woman, and 1 preferred not to identify. On average, the health service providers had worked in prevention for 6.2 (SD 5.4) years and with MSM or hijras specifically for 5.3 (SD 4.9) years.

The median age of the 8 mHealth developers who participated in individual interviews was 39 (IQR 32 to 44) years, and all were college-educated men. On average, the mHealth developers had worked in mHealth for 4.0 (SD 2.9) years and had worked in India on average for 3.3 (SD 2.9) years. All the developer participants believed that mHealth interventions could be useful, although they acknowledged that developing such a tool for the Indian context would be challenging due to differences in languages spoken, literacy, and access to technology.

Study Procedures

Prior to data collection, a member of the research team reviewed consent forms with all participants. After consenting but prior to the start of the qualitative data collection, participants completed a brief demographic questionnaire. Four focus groups [29] were conducted with MSM who used technology to meet male sex partners. Because it was difficult to identify hijras who use technology to meet sex partners—perhaps due to differences in income between MSM and hijras—the few who were identified were interviewed individually. Topics discussed in focus groups and interviews were similar. Both began with a discussion about participants' experiences using technology to meet sex partners and concluded with a discussion about suggestions for an HIV/STI prevention mHealth intervention. Additionally, we interviewed 10 health service providers who had experience working with MSM and hijra patients. The interviews focused on the ways health service providers and their organizations might use technology-based interventions to address the health needs of MSM and hijras. Each focus group lasted approximately 1.5 hours, and each individual interview lasted approximately 1 hour. Although focus groups and interviews were conducted in the preferred language of participants (Marathi, Hindi, or English), most focus groups and interviews occurred in English. During the focus groups, participants often switched between English and either Hindi or Marathi. Copies of the semistructured focus group and interview protocols are available upon request from the corresponding author.

Data Analysis

All focus groups and interviews were audiorecorded and transcribed. When participants spoke Hindi or Marathi, the recordings were translated into English by staff at The Humsafar Trust. The first and third authors, who are fluent in Marathi, Hindi, and English, compared the translated transcripts for accuracy. Because many focus group and interview participants spoke in English at least some of the time, the words of participants were retained for analysis. The English portions of the transcripts were not altered to adhere to English grammar rules. The English transcripts were entered into NVivo 10 (QSR International Pty Ltd) for content analysis [30]. Independently, the first and second authors coded all transcripts for distinct themes based on the meaning of words or phrases and then came together to compare codes and arrive at a common coding taxonomy by examining codes for frequency, strength, and relationship. The taxonomy was presented to and vetted by the entire research team to arrive at a final taxonomy, which was used to code all transcripts.

Results

Three overarching themes emerged from the data that were relevant to the development of an mHealth HIV prevention intervention for Indian MSM: content, interface, and retention features of the intervention.

Content

Content recommendations for an mHealth intervention included providing sex education for youth, providing information about STIs, and providing information and social support for PLHA. Because sex education is not uniformly taught in schools, participants suggested mHealth could provide missing sex education. Topics such as gender, identity, and sexuality as well as HIV/STI facts, proper condom use, and condom use negotiation skills were identified as being the most salient.

> *The way sex should be done, condom should be used and it should be done at safe place, all this should be shown.* [Hijra]

Participants also expressed a need for an intervention to address reasons why younger MSM and hijras might engage in condomless anal sex (barebacking), including possible coercion by older partners and perceived HIV status of a potential sex partner based on their appearance.

> *When I was 17, people used to persuade me to have bareback sex. Thankfully, I was aware of not having sex but they used to say, "Do it, nothing [bad] would happen." So, the young guys would want to try out everything so that they might just fall for it.* [MSM]
>
> *If you smell good, wear good perfume, have nice underwear, and if you are physically clean then you are considered to be [HIV negative].* [MSM]

Participants viewed an mHealth intervention as a tool to provide accurate information on STIs. MSM and hijra participants recommended that the intervention have adequate information on STIs to aid in the identification of STIs prior to having sex with a potential partner. The health service providers shared this view and acknowledged that an mHealth intervention could also bridge gaps in understanding that health service providers have about STIs. Health service providers expected an mHealth intervention to promote a fear-based approach to STIs with graphic pictures of infections to scare the community members into practicing safer sex; however, MSM and hijras expressed an aversion to this approach. Thus, while there was agreement about the need for information about STIs, health service providers and MSM and hijras had divergent views on how best to deliver the information.

MSM, hijras, and health service providers viewed an mHealth intervention as a potential source of information and support for PLHA. Topics included ways to improve positive living, including medication adherence, how to contact a health care worker in cases of emergency hospitalizations, information about health centers offering various services for PLHA, and an online support group for respondents. Participants suggested that an mHealth intervention could connect individuals with health care workers who could provide support to PLHA in case of emergency as well as routine hospitalizations and with other persons living with the virus.

> *Sometimes PLHA and MSM community people [are] hospitalized and they don't have their family members with them. At that time, there is a major problem with their care... so if we can do it by app...we can get free or voluntary caretakers for them. This would be the most important support for PLHA community.* [Health service provider]

Participants felt community support was needed for PLHA because of the stigma and discrimination associated with living with HIV. An online community would ensure privacy.

Interface

Interface recommendations for an mHealth intervention included presenting content using pictures; using videos to present stories of role models; using push notifications for testing, appointment, and medication reminders; using geolocation to link to just-in-time (JIT) services; and using telemedicine to increase access to health care providers and community services.

Nearly all participants agreed that mHealth interventions should have minimal text if MSM and hijras were to use it regularly. They suggested presenting information in pictorial formats. Some participants' expressed concerns about text-heavy content were potential user short attention spans and limited literacy, particularly among members of the hijra community.

> *Normally in our community, so many uneducated people are there and they cannot be able to know how this app would to be used.* [Hijra]

Health service providers agreed with these observations, noting that pictorial messages can aid in recall.

> *People generally understand the messages through pictures. When he sees the picture, he will try to recall that, "Yes, I have seen it somewhere." So pictures should be used much.* [Health service provider]

Overreliance on written communication could impede understanding and utility of health messaging. Images or videos were a preferred method of messaging. Community participants expressed that an mHealth intervention could also explore areas beyond the traditional ambit of health education by using role models within the community and focusing on relationships with partners. An mHealth intervention could have role model stories from persons who have navigated stigmatizing environments to form successful social and romantic relationships with other MSM and hijras. An mHealth intervention could also provide role models of PLHA by describing the journey toward accepting one's HIV positive status and living a happy and successful life.

> *[Include] videos on how people have become comfortable having HIV. Messages from role models. Some honest interview of what stages of approval [self-acceptance] the person has gone through... Something on lines of, "I was really angry, I was revengeful and I changed."* [MSM]

Because of barriers to using SMS, push notifications when used selectively were an acceptable alternative. Because it was sometimes difficult to remember to test routinely or to remember other medical appointments, participants felt an mHealth intervention could be used to promote testing among community members by setting reminders. Reminders were perceived as being particularly useful for hijra sex workers as their working schedules often affect regular testing. For PLHA, push notifications were viewed as an acceptable method for reminders to take antiretroviral or other medications.

Participants were enthusiastic about the possibilities to access information on community events and JIT services. The most mentioned JIT services that the participants wished to see included were information on sexual and gender minority–serving nongovernmental organizations and testing centers, health care providers, counselors, legal aid, and helplines.

Some health service providers recommended telemedicine for mental health counseling and to broaden outreach, particularly to MSM and hijras who were closeted and hard to reach. One health care provider articulated the financial challenge to this approach while also highlighting the potential benefit.

> *Telemedicine broadens the outreach. [There is] considerable investment in the cost of equipment, [but] support would be counseling for mental health issues, since gay men don't find it easy to open up to everyone.* [Health service provider]

While health service providers and community participants acknowledged the infrastructure costs to begin offering telemedicine, the suggestions could allow an mHealth intervention to be accessed by a greater number of clients.

Retention

Retention recommendations for an mHealth intervention included keeping it fun, developing content in regional languages, protecting confidentiality, and linking to social networking apps. Community participants insisted that the mHealth intervention be fun; otherwise, they felt members of the community would not access it regularly. They were wary of an intervention that was too serious because they felt that the members of the community would quickly lose interest in it. Another suggestion was to employ gaming mechanics. For example, 1 participant recommended employing a reward system for engaging in health-promoting behavior (ie, using a condom after last intercourse or getting tested for HIV).

There was a strong emphasis on use of local languages. English words could be misunderstood and would make the intervention inaccessible to members of the community who are less fluent in English. Both health service providers and members of the community expressed concern about the excessive use of English. For them, it was important that intervention developers embrace the multiculturalism of India.

Participants said protecting confidentiality on an mHealth intervention would be critical. Participants highlighted the stigma that the community faces and shared instances of how stigma affects uptake of services by the community. Participants stated that confidentiality would be a key aspect influencing use of an mHealth intervention by the community. Mindful that homosexuality is still illegal and highly stigmatized in India, participants said that if the intervention looked “too gay” or included nude content then it could potentially out users as sexual or gender minorities and discourage increased use of the intervention, as a substantial proportion of the community are closeted.

> *You should just have a logo and that’s about it. Logo or the identity interface should not have a gay feel. Even the name of the app can’t be too gay. And if there is a way that a person could hide the app, saying like you have this antivirus software where when it gets camouflage with something else.* [MSM]

MSM and hijras, in particular, were in favor of having access codes to ensure that the app remains private and confidential. However, they expressed concerns about information still being visible in instances of pop-up notifications. Participants recommended that confidentiality statements be featured prominently.

An additional recommendation to make the intervention more interactive and increase retention was to incorporate social networking. Participants described using various types of social media (eg, Facebook, Twitter) to improve reach. Some participants recommended that an mHealth intervention be linked to popular instant messaging apps (eg, Whatsapp) and dating apps (eg, PlanetRomeo) to allow messages to be delivered instantly by using instant message features on these apps for sending out health messages. By linking the intervention to preexisting social networking apps, an intervention could seamlessly be incorporated into participants’ lives rather than being another app they may forget to check.

Discussion

Principal Findings

Interventions using mHealth are becoming more feasible in India because of a burgeoning market for mobile technology and use of the Internet [18,19]. Although a demographic digital divide exists with millennials from higher income brackets being more likely to be mobile phone and Internet users [31], trends suggest a rapidly expanding technological market as products become more affordable. This creates new opportunities to develop culturally relevant mobile health (mHealth) interventions for a variety of health issues including HIV prevention.

Although many of the features identified by participants were similar to those reported in the literature using data from Western countries, there were some differences, including minimal use of text, the ability to deliver content in multiple regional languages, and concerns about confidentiality due to stigma and criminalization of homosexuality. For an mHealth intervention to be used by Indian MSM and hijra communities, it needs to feel safe, private, trustworthy, and confidential. To feel credible, it needs to include numerous safeguards to ensure confidentiality of the users and not appear “too gay.” An mHealth intervention could be an avenue for strengthening care and support services to MSM and hijra PLHA. Features such as reminders could improve adherence to medication and routine health care checkups. An intervention could also provide much needed emotional and physical support to PLHA. If the app mediated mental health counseling and access to JIT services and community events, it could empower members of the community to make informed choices about their health and sexuality.

Strengths and Limitations

A strength of this study is the increased credibility resulting from triangulating data from focus groups with MSM and individual interviews with hijras, health service providers, and mHealth developers. However, data were collected only from persons living in Mumbai, so transferability to other Indian or South and Southeast Asian contexts, particularly to rural Indian contexts, is unknown. In addition, we were only able to recruit 4 hijras who had experience using technology to meet male sex partners. Similar to formative studies from Western countries, a limitation is that the focus was on preventive education, testing, and medication adherence rather than the on the entire continuum of care [27,32].

Conclusion

Although participants recommended keeping the intervention simple, they also suggested a lot of content and features. When developing an intervention, it will be important for researchers, practitioners, and developers to remain cognizant of the trade-off between developing an intervention that will achieve the desired health outcome and addressing the diverse needs of potential users of the intervention. While members of the community and health service providers have some divergent opinions on content and retention features, there is broad support for the development of an mHealth intervention to prevent HIV transmission and support persons living with the virus. Where differing views exist, such as use of pictures of graphic STIs, additional research is needed to ensure uptake of messaging. However, it should be noted that fear-based messaging techniques for HIV prevention have been shown to be ineffective [33]. Although our findings do not represent a framework for a particular intervention, practitioners and researchers developing interventions for Indian MSM and hijras should consider the wishes of the community if they decide to incorporate mobile technology into their prevention efforts.

Conflicts of Interest

None declared.

References

1. National AIDS Control Organisation: Annual Report 2015-16. URL: http://naco.gov.in/sites/default/files/Annual%20Report%202015-16_NACO.pdf [accessed 2018-04-14] [WebCite Cache ID 6wihDgxJU]
2. National AIDS Control Organisation: India HIV estimations 2015. URL: http://www.naco.gov.in/sites/default/files/India%20HIV%20Estimations%202015.pdf [accessed 2018-04-14] [WebCite Cache ID 6wihOkoWk]
3. Solomon SS, Srikrishnan AK, Sifakis F, Mehta SH, Vasudevan CK, Balakrishnan P, et al. The emerging HIV epidemic among men who have sex with men in Tamil Nadu, India: geographic diffusion and bisexual concurrency. AIDS Behav 2010 Oct;14(5):1001-1010 [FREE Full text] [doi: 10.1007/s10461-010-9711-2] [Medline: 20467890]
4. Setia MS, Lindan C, Jerajani HR, Kumta S, Ekstrand M, Mathur M, et al. Men who have sex with men and transgenders in Mumbai, India: an emerging risk group for STIs and HIV. Indian J Dermatol Venereol Leprol 2006;72(6):425-431 [FREE Full text] [Medline: 17179617]
5. Li DH, Rawat S, Rhoton J, Patankar P, Ekstrand ML, Rosser BRS, et al. Harassment and violence among men who have sex with men (MSM) and hijras after reinstatement of India sodomy law. Sex Res Soc Policy 2016 Dec 17;14(3):324-330. [doi: 10.1007/s13178-016-0270-9]
6. Mimiaga MJ, Reisner SL, Closson EF, Perry N, Perkovich B, Nguyen T, et al. Self-perceived HIV risk and the use of risk reduction strategies among men who engage in transactional sex with other men in Ho Chi Minh City, Vietnam. AIDS Care 2013 Aug;25(8):1039-1044 [FREE Full text] [doi: 10.1080/09540121.2012.748873] [Medline: 23252673]
7. Satpathy M. Emerging healthcare issues and needs of young MSM and TG in India: needs assessment from multi-stakeholders perspective. BMC Infectious Diseases 2012;12(Suppl 1):1. [doi: 10.1186/1471-2334-12-s1-p91]
8. Beratarrechea A, Lee AG, Willner J, Jahangir E, Ciapponi A, Rubinstein A. The impact of mobile health interventions on chronic disease outcomes in developing countries: a systematic review. Telemed e-Health 2014;20(1):75-82.
9. Burton J, Brook G, McSorley J, Murphy S. The utility of short message service (SMS) texts to remind patients at higher risk of STIs and HIV to reattend for testing: a controlled before and after study. Sex Transm Infect 2014 Feb;90(1):11-13. [doi: 10.1136/sextrans-2013-051228] [Medline: 24064987]
10. Chib A, Wilkin H, Ling LX, Hoefman B, Van Biejma H. You have an important message! Evaluating the effectiveness of a text message HIV/AIDS campaign in Northwest Uganda. J Health Commun 2012;17 Suppl 1:146-157. [doi: 10.1080/10810730.2011.649104] [Medline: 22548607]
11. Curioso WH, Quistberg DA, Cabello R, Gozzer E, Garcia PJ, Holmes KK, et al. It's time for your life: how should we remind patients to take medicines using short text messages? AMIA Annu Symp Proc 2009;2009:129-133 [FREE Full text] [Medline: 21633523]
12. da Costa TM, Barbosa BJP, Gomes ECDA, Sigulem D, de Fatima Marin H, Filho AC, et al. Results of a randomized controlled trial to assess the effects of a mobile SMS-based intervention on treatment adherence in HIV/AIDS-infected Brazilian women and impressions and satisfaction with respect to incoming messages. Int J Med Inform 2012 Apr;81(4):257-269 [FREE Full text] [doi: 10.1016/j.ijmedinf.2011.10.002] [Medline: 22296762]
13. Dowshen N, Kuhns LM, Gray C, Lee S, Garofalo R. Feasibility of interactive text message response (ITR) as a novel, real-time measure of adherence to antiretroviral therapy for HIV+ youth. AIDS Behav 2013 Jul;17(6):2237-2243. [doi: 10.1007/s10461-013-0464-6] [Medline: 23546844]
14. Furberg RD, Uhrig JD, Bann CM, Lewis MA, Harris JL, Williams P, et al. Technical implementation of a multi-component, text message-based intervention for persons living with HIV. JMIR Res Protoc 2012 Nov 16;1(2):e17 [FREE Full text] [doi: 10.2196/resprot.2017] [Medline: 23612237]
15. Telecom Regulatory Authority of India. Indian Telecom services performance indicators: April-June 2009 URL: http://www.trai.gov.in/Default.asp [WebCite Cache ID 6wigGbd15]
16. Rodrigues R, Shet A, Antony J, Sidney K, Arumugam K, Krishnamurthy S, et al. Supporting adherence to antiretroviral therapy with mobile phone reminders: results from a cohort in South India. PLoS One 2012 Aug;7(8):e40723 [FREE Full text] [doi: 10.1371/journal.pone.0040723] [Medline: 22952574]
17. Shet A, Costa AD, Kumarasamy N, Rodrigues R, Rewari BB, Ashorn P, et al. Effect of mobile telephone reminders on treatment outcome in HIV: evidence from a randomised controlled trial in India. BMJ 2014 Oct 24;347(aug06 2):g5978-g5978. [doi: 10.1136/bmj.g5978]
18. Zenith. Smartphone penetration to reach 66% in 2018 URL: https://www.zenithmedia.com/smartphone-penetration-reach-66-2018/ [WebCite Cache ID 6xJJYj6Vs]
19. International Telecommunication Union. ITU ICT facts and figures 2016 URL: https://www.itu.int/en/ITU-D/Statistics/Documents/facts/ICTFactsFigures2016.pdf [WebCite Cache ID 6wigi98Rt]
20. Cordova D, Bauermeister JA, Fessler K, Delva J, Nelson A, Nurenberg R, et al. A community-engaged approach to developing an mHealth HIV/STI and drug abuse preventive intervention for primary care: a qualitative study. JMIR Mhealth Uhealth 2015 Dec;3(4):e106 [FREE Full text] [doi: 10.2196/mhealth.4620] [Medline: 26685288]
21. Jennings L, Gagliardi L. Influence of mHealth interventions on gender relations in developing countries: a systematic literature review. Int J Equity Health 2013 Oct 16;12:85-104 [FREE Full text] [doi: 10.1186/1475-9276-12-85] [Medline: 24131553]

22. Horvath KJ, Alemu D, Danh T, Baker JV, Carrico AW. Creating effective mobile phone apps to optimize antiretroviral therapy adherence: perspectives from stimulant-using HIV positive men who have sex with men. JMIR Mhealth Uhealth 2016 Apr 15;4(2):e48 [FREE Full text] [doi: 10.2196/mhealth.5287] [Medline: 27084049]
23. Kirk GD, Himelhoch SS, Westergaard RP, Beckwith CG. Using mobile health technology to improve HIV care for persons living with HIV and substance abuse. AIDS Res Treat 2013;2013:194613 [FREE Full text] [doi: 10.1155/2013/194613] [Medline: 24381751]
24. Grandin M. Improving therapeutic success through personalised mobile health tools. Int J Integrated Care 2014;14:90-91.
25. Hightow-Weidman LB, Fowler B, Kibe J, McCoy R, Pike E, Calabria M, et al. HealthMpowerment.org: development of a theory-based HIV/STI website for young black MSM. AIDS Educ Prev 2011 Feb;23(1):1-12 [FREE Full text] [doi: 10.1521/aeap.2011.23.1.1] [Medline: 21341956]
26. Muessig KE, Nekkanti M, Bauermeister J, Bull S, Hightow-Weidman LB. A systematic review of recent smartphone, Internet and Web 2.0 interventions to address the HIV continuum of care. Curr HIV/AIDS Rep 2015 Jan 28;12(1):173-190. [doi: 10.1007/s11904-014-0239-3] [Medline: 25626718]
27. Sullivan PS, Jones J, Kishore N, Stephenson R. The roles of technology in primary HIV prevention for men who have sex with men. Curr HIV/AIDS Rep 2015 Dec;12(4):481-488. [doi: 10.1007/s11904-015-0293-5] [Medline: 26519083]
28. Wilkerson JM, Patankar P, Rawat SM, Simon RBR, Shukla KM, Rhoton J, et al. Recruitment strategies of Indian men who have sex with men in the state of Maharashtra into an online survey. Int J Sex Health 2016;28(3):221-227 [FREE Full text] [doi: 10.1080/19317611.2016.1193079] [Medline: 27668029]
29. Krueger R, Casey M. Focus Groups: A Practical Guide for Applied Research. 4th edition. Thousand Oaks: Sage Publications, Inc; 2009.
30. Weber R. Basic Content Analysis. 2nd edition. Newburry Park: Sage Publications; 1990.
31. Pew Research Center. Smartphone ownership and Internet usage continues to climb in emerging economies URL: http://assets.pewresearch.org/wp-content/uploads/sites/2/2016/02/pew_research_center_global_technology_report_final_february_22__2016.pdf [WebCite Cache ID 6wigtrM8g]
32. Simoni JM, Kutner BA, Horvath KJ. Opportunities and challenges of digital technology for HIV treatment and prevention. Curr HIV/AIDS Rep 2015 Sep 28;12(4):437-440. [doi: 10.1007/s11904-015-0289-1] [Medline: 26412082]
33. Albarracín D, Gillette JC, Earl AN, Glasman LR, Durantini MR, Ho M. A test of major assumptions about behavior change: a comprehensive look at the effects of passive and active HIV-prevention interventions since the beginning of the epidemic. Psychol Bull 2005 Nov;131(6):856-897 [FREE Full text] [doi: 10.1037/0033-2909.131.6.856] [Medline: 16351327]

Abbreviations

IQR: interquartile range
JIT: just-in-time
MSM: men who have sex with men
PLHA: persons living with HIV/AIDS
SMS: short message service
STI: sexually transmitted infection

Edited by G Eysenbach; submitted 29.09.17; peer-reviewed by J Simoni, I Fakoya, M Downing, S Aggarwal; comments to author 03.12.17; revised version received 17.02.18; accepted 17.03.18; published 03.05.18.

Please cite as:
Rawat S, Wilkerson JM, Lawler SM, Patankar P, Rosser BRS, Shukla K, Butame S, Ekstrand ML
Recommendations for the Development of a Mobile HIV Prevention Intervention for Men Who Have Sex With Men and Hijras in Mumbai: Qualitative Study
JMIR Public Health Surveill 2018;4(2):e46
URL: http://publichealth.jmir.org/2018/2/e46/
doi:10.2196/publichealth.9088
PMID:29724705

Original Paper

Trends in HIV Terminology: Text Mining and Data Visualization Assessment of International AIDS Conference Abstracts Over 25 Years

Nicole Dancy-Scott[1*], MPH; Gale A Dutcher[1*], MS, MLS; Alla Keselman[1*], MA, PhD; Colette Hochstein[1*], MLS, DMD; Christina Copty[2*], MPH (Epidemiology); Diane Ben-Senia[2*], MPH; Sampada Rajan[2*], MEM; Maria Guadalupe Asencio[2*], MS; Jason Jongwon Choi[2*], MS

[1]Division of Specialized Information Services, National Library of Medicine, Bethesda, MD, United States

[2]ICF, Fairfax, VA, United States

[*]all authors contributed equally

Corresponding Author:
Nicole Dancy-Scott, MPH
Division of Specialized Information Services
National Library of Medicine
6707 Democracy Boulevard, Suite 510
Bethesda, MD,
United States
Phone: 1 301 496 3147
Fax: 1 301 480 3537
Email: nicole.scott@nih.gov

Abstract

Background: The language encompassing health conditions can also influence behaviors that affect health outcomes. Few published quantitative studies have been conducted that evaluate HIV-related terminology changes over time. To expand this research, this study included an analysis of a dataset of abstracts presented at the International AIDS Conference (IAC) from 1989 to 2014. These abstracts reflect the global response to HIV over 25 years. Two powerful methodologies were used to evaluate the dataset: text mining to convert the unstructured information into structured data for analysis and data visualization to represent the data visually to assess trends.

Objective: The purpose of this project was to evaluate the evolving use of HIV-related language in abstracts presented at the IAC from 1989 to 2014.

Methods: Over 80,000 abstracts were obtained from the International AIDS Society and imported into a Microsoft SQL Server database for data processing and text mining analyses. A text mining module within the KNIME Analytics Platform, an open source software, was then used to mine the partially processed data to create a terminology corpus of key HIV terms. Subject matter experts grouped the terms into categories. Tableau, a data visualization software, was used to visualize the frequency metrics associated with the terms as line graphs and word clouds. The visualized dashboards were reviewed to discern changes in terminology use across IAC years.

Results: The major findings identify trends in HIV-related terminology over 25 years. The term “AIDS epidemic” was dominantly used from 1989 to 1991 and then declined in use. In contrast, use of the term “HIV epidemic” increased through 2014. Beginning in the mid-1990s, the term “treatment experienced” appeared with increasing frequency in the abstracts. Use of terms identifying individuals as “carriers or victims” of HIV rarely appeared after 2008. Use of the terms “HIV positive” and “HIV infected” peaked in the early-1990s and then declined in use. The terms “men who have sex with men” and “MSM” were rarely used until 1994; subsequently, use of these terms increased through 2014. The term “sex worker” steadily increased in frequency throughout conference years, whereas the term “prostitute” decreased over time.

Conclusions: The results of this study highlight changes in HIV terminology use over 25 years, including the addition, disappearance, and changing use of terms that reflect advances in HIV research and medical practice and destigmatization of the disease. Coupled with findings from related quantitative research, HIV-related terminology recommendations based on results of this study are included. Adoption of these recommendations will further efforts to use less stigmatizing language and facilitate effective communication between health professionals and people affected by HIV.

(JMIR Public Health Surveill 2018;4(2):e50) doi:10.2196/publichealth.8552

KEYWORDS

acquired immunodeficiency syndrome; data mining; history; HIV infections; terminology

Introduction

Background

Since the first AIDS cases were reported in the United States in June 1981 [1], the vocabulary of HIV has evolved with advances in HIV research and medical practice. Text mining and visualization provide a method to investigate this evolution of HIV terminology. Using this combination strategy, the cultural and linguistic trajectory of HIV-related societal discourse can be characterized more precisely than possible with less structured approaches.

The relationship between progress to understand and treat diseases and the surrounding conversation is bidirectional. This mutual influence extends to all spheres of communication, including between patient and provider, and can influence behaviors and affect health outcomes.

For example, the term “AIDS carrier” was used often at the onset of the HIV epidemic. With the discovery of the virus that causes AIDS, the term was recognized as both biologically incorrect and inherently stigmatizing, potentially creating and reinforcing negative stereotypes of people living with HIV [2]. Kelly et al evaluated the language used to refer to people with substance use disorders. This research showed that “The language used to describe health conditions reflects and influences our attitudes and approaches to addressing them, even to the extent of suggesting that a health condition is a moral, social, or criminal issue” [3]. Stigmatizing language that assigns blame can deter individuals from seeking care and even affect clinicians’ treatment decisions. When language helps people to feel empowered, they are more likely to seek care [3].

In addition, language often reflects sociocultural and historical trends over time [4]. Michel et al created a corpus containing more than 5 million digitized books—4% of all books ever published—to quantitatively investigate cultural trends. This research demonstrated that computational analysis of digitized texts (*culturomics*) offers insights across diverse fields, including the evolution of grammar, adoption of technology, and historical epidemiology. The research also showed that such analysis could reveal the concepts of great societal interest across time and the associated linguistic changes (eg, the shifting use of “The Great War” and “World War I” to discuss the same event) [4]. Funk, in a frequency analysis of a corpus of journal articles exploring changes in the medical library profession over 50 years, revealed changing trends in the profession, such as the growing preference for evidence-based information and an increasing focus on health in addition to medicine [5].

To the best of our knowledge informed by a thorough literature search, no published quantitative studies document changes in HIV-related terminology over time. The few published works investigating HIV terminologies are qualitative and focus on word use in a specific context. For example, Mupenda et al interviewed people living with HIV in the Democratic Republic of the Congo about the labels applied to them in their community and social groups. Using text mining software to code and analyze the information, the study identified mostly derogatory labels (eg, “a disease that one chooses to get,” “a sick hen,” and “walking or standing corpse”) [6]. The study, however, did not explore how the labeling may have changed over time. Duby et al interviewed women in Africa who had previously participated in an HIV prevention study to better understand how participants interpret questions related to sexual behaviors. The study results highlighted how cultural biases, misunderstanding of research terms, and incorrect or absent translations for vernacular expressions can negatively affect HIV research [7]. Several nonresearch publications also make prescriptive recommendations about preferred terminology in clinical, community, and research contexts. For example, an annotation of terminology recommendations developed by advocates and activists living with HIV recommends using language that puts people first, avoiding the word “infection” unless necessary, and not merging HIV and AIDS into a single term [8].

The Study

To expand the research on HIV terminology into a quantitative, temporal characterization of HIV terminology, a digitized dataset of abstracts presented at the International AIDS Conference (IAC) from 1989 to 2014 was analyzed. The IAC abstracts were used for this study because they were available from a sole source, the International AIDS Society. In addition, the IAC, initially convoked in 1985, represents the global response to HIV [9]. Analysis of the IAC dataset involved a hybrid process that integrated automated text mining and visualization with manual subject matter expert reviews. The text mining analysis incorporated algorithms tested successful in similar research to determine word relevancy and for term extraction [10,11]. Data processing for this study was conducted using the KNIME Analytics Platform, which is an open-source software used widely in the academic community for text mining [12]. The KNIME Analytics Platform includes modules for data blending, data and text mining, and machine learning, as well as other technical components. This study used the KNIME text mining module that is designed to perform statistical analysis on textual data (represented in a term-document frequency matrix) for further data analysis [13].

Results of the text mining analyses were visualized using Tableau Desktop to identify and highlight trends in HIV-related terminology use across the conference years. Tableau is a business intelligence interactive data visualization and dashboard tool [14]. These results can provide some insight and help generate hypotheses about the evolving historical and sociocultural context of the HIV epidemic.

Study Objectives

The rationale for this study was multifold. First, we wanted to investigate using automated text analysis to identify terminology trends in a health domain, specifically HIV. We also aimed to construct a digitized corpus representing the evolution of HIV terminology in scientific literature. In addition to its value as a historical document, such a representation may provide a foundation for follow-up studies on the correspondence between social and linguistic changes, the uniformity of linguistic changes, and the implementation of terminology guidelines. It may also be useful to include the results of this study in guidelines and training materials to illustrate the relationship between language and stigmatization of individuals living with a disease.

Although the study was largely exploratory, results were expected to include terminology changes reflecting (1) Reduced stigmatization of individuals living with HIV, and (2) New strategies to treat HIV and reduce HIV transmission.

The following research questions were investigated in this study: (1) What changes in HIV-related language, if any, can be determined from IAC abstracts over 25 years?; (2) Are there any HIV-related terms that appear in abstracts only in specific years?; and (3) How does the HIV-related language used in the abstracts reflect the history of the HIV epidemic during the same 25-year period?

Methods

Project Workflow

The study progressed in four stages as shown in Figure 1: data source development, data processing, terminology corpus creation, and visualization and analysis.

Data Source Development

A total of 88,922 abstracts were obtained from the International AIDS Society for IAC conference years 1989 to 2014. The number of abstracts per IAC conference year varied, ranging from 5547 in 1989, 10,193 in 2006, to 2335 in 2014. The extensive dataset was delivered in comma-separated values files. The comma-separated values files were imported into an Excel (Microsoft) spreadsheet for data cleaning, a step necessary to convert the machine readable data into a more appropriate dataset for text mining. The data cleaning process included eliminating duplicate abstracts and removing HTML characters. This process did not change any words or alter the contents of the abstracts. The cleaned-up data were then imported into a Microsoft SQL Server relational database. The data from the Microsoft SQL Server database were converted to text files for data processing and text mining analysis.

Data Processing

Using the KNIME software, the IAC abstracts were parsed to extract and represent the abstract text as a data structure. The parsed text was represented by the inherent KNIME data types called “DocumentCell” and “DocumentValue.” The parsed text was collected and saved as a data object or vector in which each element is an abstract. This processing allowed the application to manipulate the abstract text for further analysis, specifically to identify terms of interest and estimate relevant metrics of frequency. The KNIME software was then used to mine the partially processed data to create a terminology corpus of key HIV terms for analysis. The process began by preprocessing the data object of abstracts to filter out irrelevant terms (eg, stop words [the, a, an], numbers, punctuation, and diacritics [umlauts, accents]).

Terminology Corpus Creation

The process to create the terminology corpus continued with the use of the KNIME Predictive Analytics software to separate the abstracts by conference year into unigrams. The number of words per abstract per conference year ranged from 36,000 to 60,000. Each word was associated with a KNIME-calculated metric of document frequency (the number of abstracts in which a term appears in a conference year).

The unigrams were exported to an Excel (Microsoft) spreadsheet and organized in ascending order of document frequency. A visual assessment of the words and their associated document frequencies showed that the majority of the terms had a document frequency of over 100. An HIV subject matter expert reviewed the spreadsheet to identify an initial set of unigrams (eg, “HIV,” “AIDS,” and “therapy”) with a document frequency of 100 or higher to investigate further. The subject matter expert also scanned the spreadsheet for any unigrams below the frequency threshold but judged relevant based on her HIV-related expertise. Given the scope of the project and the data import limitations of the free Tableau Desktop visualization software used, the subject matter expert limited this initial set of unigrams to 15 to 20 key unigrams. A second HIV subject matter expert reviewed the set of key unigrams to confirm their HIV-related relevancy.

KNIME Predictive Analytics software was again used to generate related bigrams (two-word terms) and trigrams (three-word terms) from the key unigrams. For example, the bigram “antiretroviral therapy” and the trigram “combination antiretroviral therapy” were generated from the unigram “therapy.” From the set of 15 to 20 key unigrams, the KNIME software generated approximately 15,000 bigrams and trigrams.

Again, using an initial document frequency threshold of 100, two HIV subject matter experts reviewed the unigrams, bigrams, and trigrams to identify key 1-, 2-, 3-, and 4-word terms to create the HIV terminology corpus. Both subject matter experts completed this step independently and then compared the results to ensure that all possible terms were included in the corpus. Drawing on their knowledge of HIV, the subject matter experts also included some unigrams, bigrams, and trigrams with a document frequency of less than 100 (eg, “drug holiday”) and some 4-grams (eg, “people who inject drugs” and “highly active antiretroviral therapy”) to the terminology corpus. The final set of HIV-related terms (the terminology corpus) included 1-, 2-, 3-, and 4-word terms.

The objective of this research was to evaluate the terms used in the IAC abstracts. Hence, synonyms were not normalized, and the words used in the abstracts were retained in the terminology corpus unchanged, except for the normalizations mentioned below. Research investigating use of Term Frequency-Inverse Document Frequency (TF-IDF) for query retrieval noted that the algorithm does not equate a word with its plural and that this limitation could potentially produce inaccuracies [10]. To overcome this limitation of the TF-IDF calculation, the terminology corpus creation process used for this study accounted for plural terms. In addition, terms with symbols were normalized by converting the symbols to equivalent terms. For example, *HIV+* was converted to “HIV positive.” Related terms, abbreviations, and acronyms for terms (eg, “ART” for “antiretroviral therapy”) were also added to the terminology corpus. Any acronym representing more than one word or phrase (eg, “STI,” the acronym for both “structured treatment interruption” and “sexually transmitted infection”) was not included in the terminology corpus. Once the terminology corpus was finalized, related terms and their abbreviations and acronyms were grouped into categories (eg, “HIV prevention” and “living with HIV”). The terms included in each category were mutually exclusive, and each term was classified into one category.

Given the extent of the IAC dataset (80,000 abstracts), it was impossible to verify that the meaning or contextual use of terms in the terminology corpus was consistent across all conference years. However, the validity of the terminology corpus was confirmed in a subset of abstracts (see Multimedia Appendix 1 for the final terminology corpus).

Metrics Evaluated

Two frequency metrics were calculated to identify changes in HIV terminology use within and across conference years: relative term frequency (relTF) and TF-IDF. To derive these metrics, the following additional frequencies were calculated: absolute TF, TF, document frequency, IDF, and moving average of relTF.

Relative Term Frequency

RelTF is a normalized metric used to compare HIV terminology within and across conference years. To calculate relTF, absolute TF was first estimated by conference year. Absolute TF represents the total number of times that a term appears in all abstracts in a conference year. For example, if conference year 1 includes three abstracts and the term “AIDS patient” appears 5 times in abstract 1, 10 times in abstract 2, and 15 times in abstract 3, the absolute frequency for “AIDS patient” in year 1 is 30.

TF represents absolute TF normalized by the number of *all* words appearing in all abstracts in a conference year. The total number of *all* words is called *the vocabulary size*. TF is a normalized value used to compare HIV terminology use within and across conference years.

TF within a conference year was calculated as shown in equation 1, where T is a specific term, and Y is a specific conference year:

$$TF(T,Y)=(absolute\ term\ frequency)/(document\ corpus) \quad (1)$$

Table 1 provides an example contrasting absolute term frequencies and term frequencies by conference year for two terms.

RelTF represents TF multiplied by 1,000,000. Thus, relative TF is the number of times in which a term is expected to appear in a document of 1,000,000 words.

Figure 1. Project workflow. IAC: International AIDS Conference.

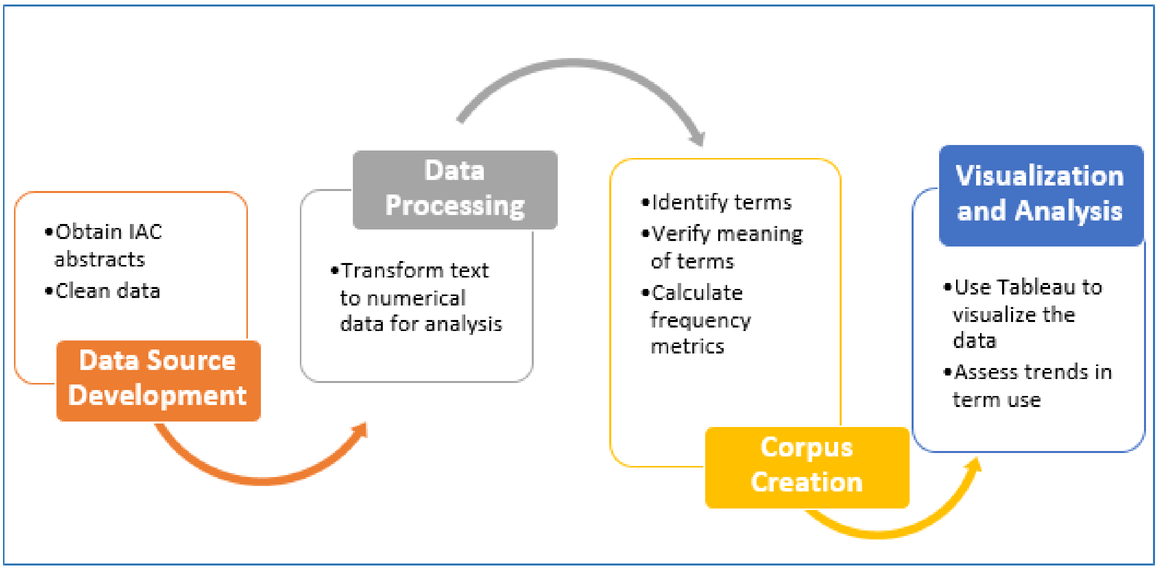


Table 1. Absolute term frequency and term frequency for select terms.

Term	IAC[a] conference year	Vocabulary size	Absolute term frequency	Term frequency
Affected communities	1989	337,263	1	0.00000242
Affected communities	1990	213,980	2	0.00000775
Affected communities	1991	239,047	2	0.00000695
Affected communities	2010	653,819	29	0.00004110
Affected communities	2012	413,162	9	0.00002039
Affected communities	2014	269,633	10	0.00003399
AIDS carrier	1989	337,263	7	0.00001693
AIDS carrier	1990	213,980	1	0.00000388
AIDS carrier	1993	355,524	2	0.00000478
AIDS carrier	1994	240,841	4	0.00001424
AIDS carrier	1998	559,343	3	0.00000499
AIDS carrier	2002	845,289	2	0.00000218
AIDS carrier	2004	866,840	2	0.00000216

[a] IAC: International AIDS Conference.

Term Frequency-Inverse Document Frequency

TF-IDF is a normalized metric that reflects the contextual importance of a term in abstracts for a specific conference year. To calculate TF-IDF, document frequency and IDF were calculated first. Abstract document frequency represents the number of abstracts in which a term appears in a conference year. At this level, each abstract in a conference year was considered independently. Document frequency at the abstract level is used to compare HIV terminology use within but not across conference years. Conference year document frequency represents the number of conference years in which a term appears in abstracts. Accordingly, all abstracts in a conference year including a term were combined and considered as a single document. Thus, at this level, document frequency ranged from 1 to 16, the number of conference years represented in the IAC dataset. The document frequencies by conference year were used as interim metrics to calculate IDF.

The frequency with which a term is used may not reflect its topical importance. For example, across the IAC abstracts, the conjunction “and” appears much more frequently than any HIV-related term. IDF gives more weight to key, less frequently used terms and less weight to less relevant, more frequently used terms. The intent is to reduce the absolute TF by a factor that grows with its TF. For this study, IDF was estimated by conference years.

IDF was calculated as shown in equation 2, where T is a specific term, N is the number of conference years, and $DF(T)$ is the number of conference years that include abstracts in which a specific term appears:

$$(2)\quad IDF(T)=\log(1+N/(DFT)))$$

TF-IDF represents the product of TF (absolute TF normalized by the total number of words in all abstracts in a conference year) and IDF (the number of conference years that include abstracts in which the term appears).

TF-IDF for a conference year is calculated by multiplying TF by IDF as shown in equation 3, where absolute frequency (T,Y) is the number of times term T appears in conference year Y, vocabulary size (Y) is the number of terms that appear in abstracts in conference year Y, $N=16$ is the number of conference years represented in the IAC dataset, and document frequency (Y) is the number of conference years in which term T appears in at least one abstract:

$$(3)\quad TFIDF(T,Y)=TF*IDF=[(absolute\ frequency\ (T,Y))/(vocabulary\ size\ document\ corpus\ (Y))]*[log(1+N/(document\ frequency(T)))]$$

Moving Average of Relative Term Frequency

A moving average is used to *smooth out* random variations across time to make trends more evident. Moving averages are calculated for overlapping time spans rather than for discrete, mutually exclusive periods of time.

For this project, the moving average of relTF was estimated for overlapping conference years as shown in equation 4, where t is the conference year:

$$(4)\quad Mean\ Average(t)=(X(t-2)+X(t-1)+X(t))/3$$

Data Visualization

Once the terminology corpus of key HIV-related terms and their associated frequency metrics were calculated, Tableau was used to extract the data from the relational database and display the information in dashboards. The dashboard displays included line graphs, tables, and word clouds. A panel that included HIV public health subject matter experts, medical librarians, evaluation specialists, and technical experts in data analysis reviewed the dashboards to identify changes in HIV terminology across IAC conference years.

Results

Analysis by Terminology Category

The following results include both line graphs and word clouds and are grouped into two categories: people descriptors and treatment descriptors.

People Descriptors

Figure 2 shows the relTFs for terms in the category “living with HIV.” Use of terms identifying individuals as “carriers” or “victims” of HIV (eg, “AIDS carrier,” “AIDS victim,” “HIV carrier,” and “HIV victim”) gradually disappeared from IAC abstracts, rarely appearing after 2008. Use of the terms “HIV positive” and “HIV infected” peaked in the early-1990s and then declined as evidenced by their decreasing relTFs.

Figure 3 is a line graph displaying the relTFs for the terms “gay” and “homosexual” included in the category “homosexual.” Use of both terms declined through conference year 2014.

Figure 4 is a line graph showing the relTFs for the terms included in the terminology category “MSM.” The terms “men who have sex with men” and “MSM” were rarely used until 1994; subsequent use of these terms increased through 2014.

Figure 5 shows a line graph of the relTfs for the terms included in the terminology category “sex worker.” Across the conference years, the most notable contrast is the declining use of the term “prostitute” and the increasing use of the term “sex worker.”

Figures 6 and 7 show word clouds representing the TF-IDFs for terms included in the category “alcohol and drug use” for conference years 1989 and 2014, respectively. Word clouds were used to highlight results of this category because of the dramatic shift noted between terms used in 1989 and 2014 that are more evident in the word cloud than a line graph. The word clouds illustrate that the use of the terms “drug addict,” “drug abuser,” and “drug abuse” in 1989 were replaced by the terms “people who use drugs” and “people who inject drugs” by 2014.

Figure 8 shows a line graph of the relTFs for the terms included in the terminology category “older adults.” Starting in the early 1990s, the term “seniors” was used more frequently that the term “older patient.”

Figure 2. Relative term frequencies for terms associated with the category “living with HIV.”.

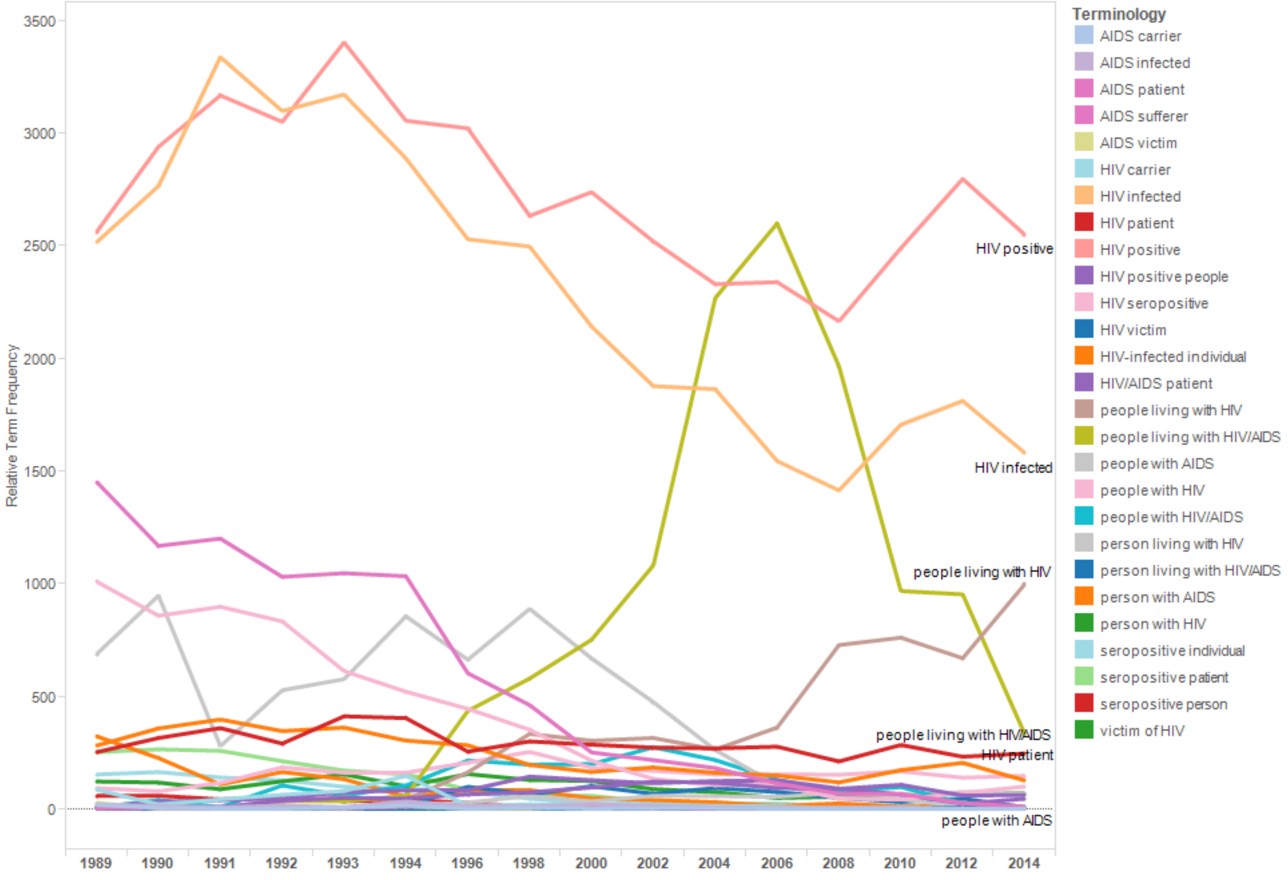


Figure 3. Relative term frequencies for terms associated with the category “homosexual.”.

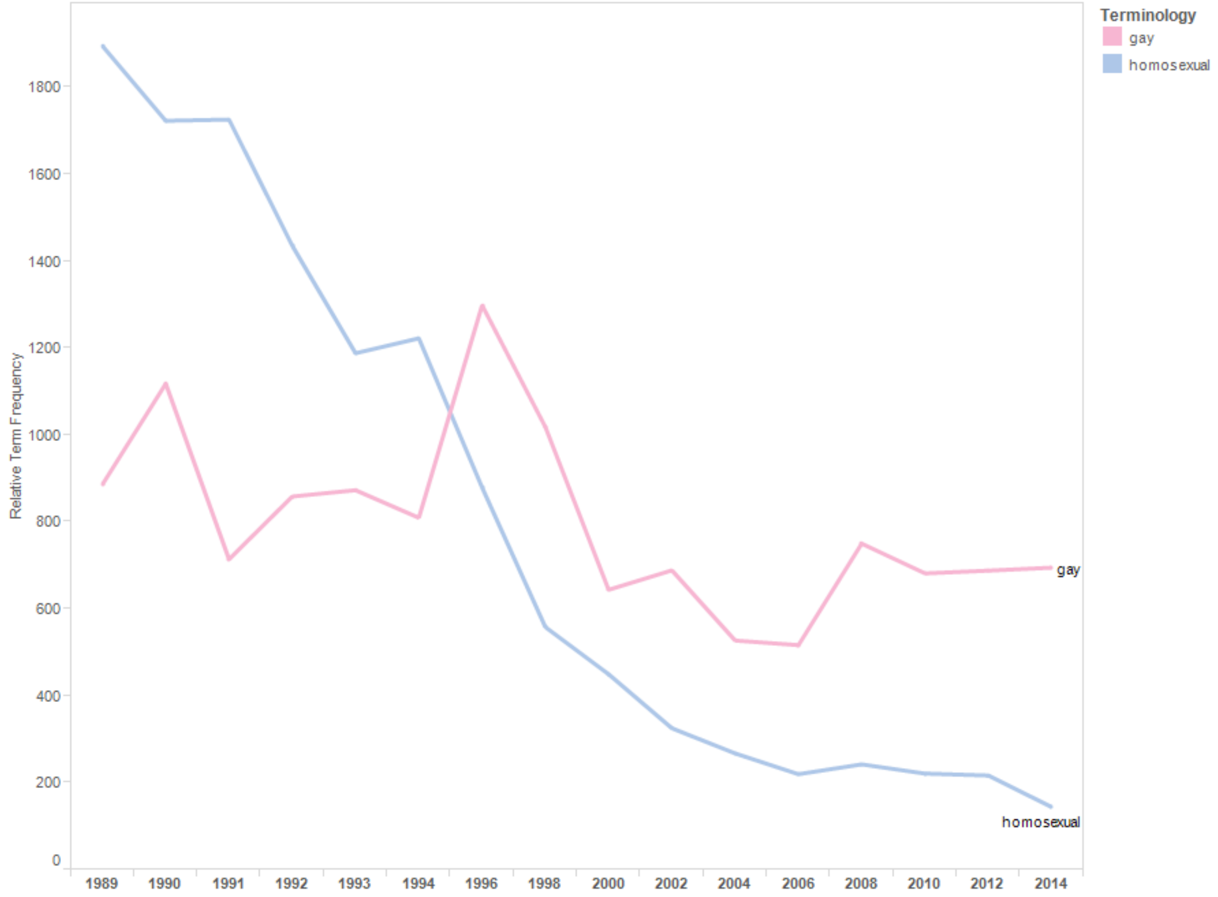


Figure 4. Relative term frequencies for terms associated with the category “MSM.”.

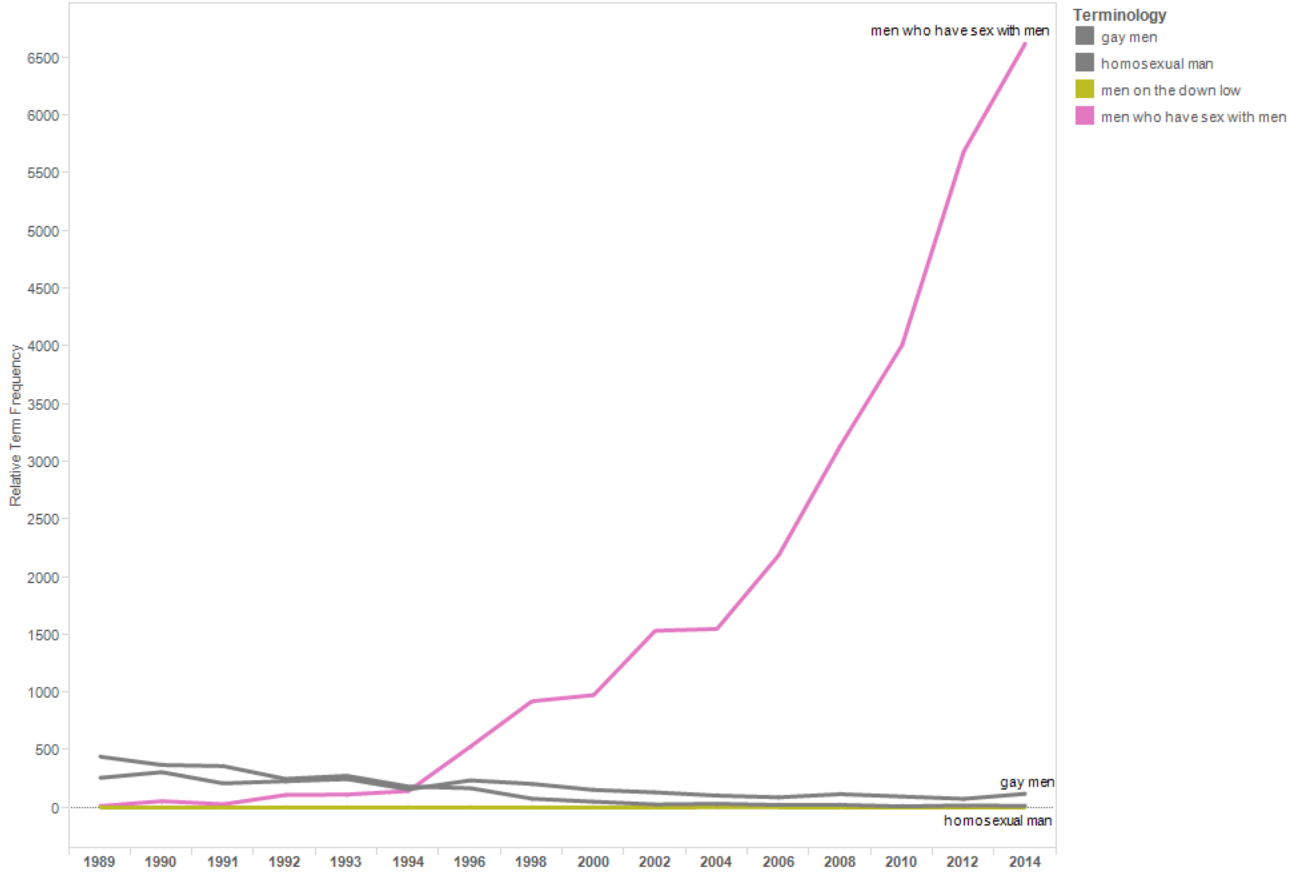


Figure 5. Relative term frequencies for terms associated with the category “sex worker.”.

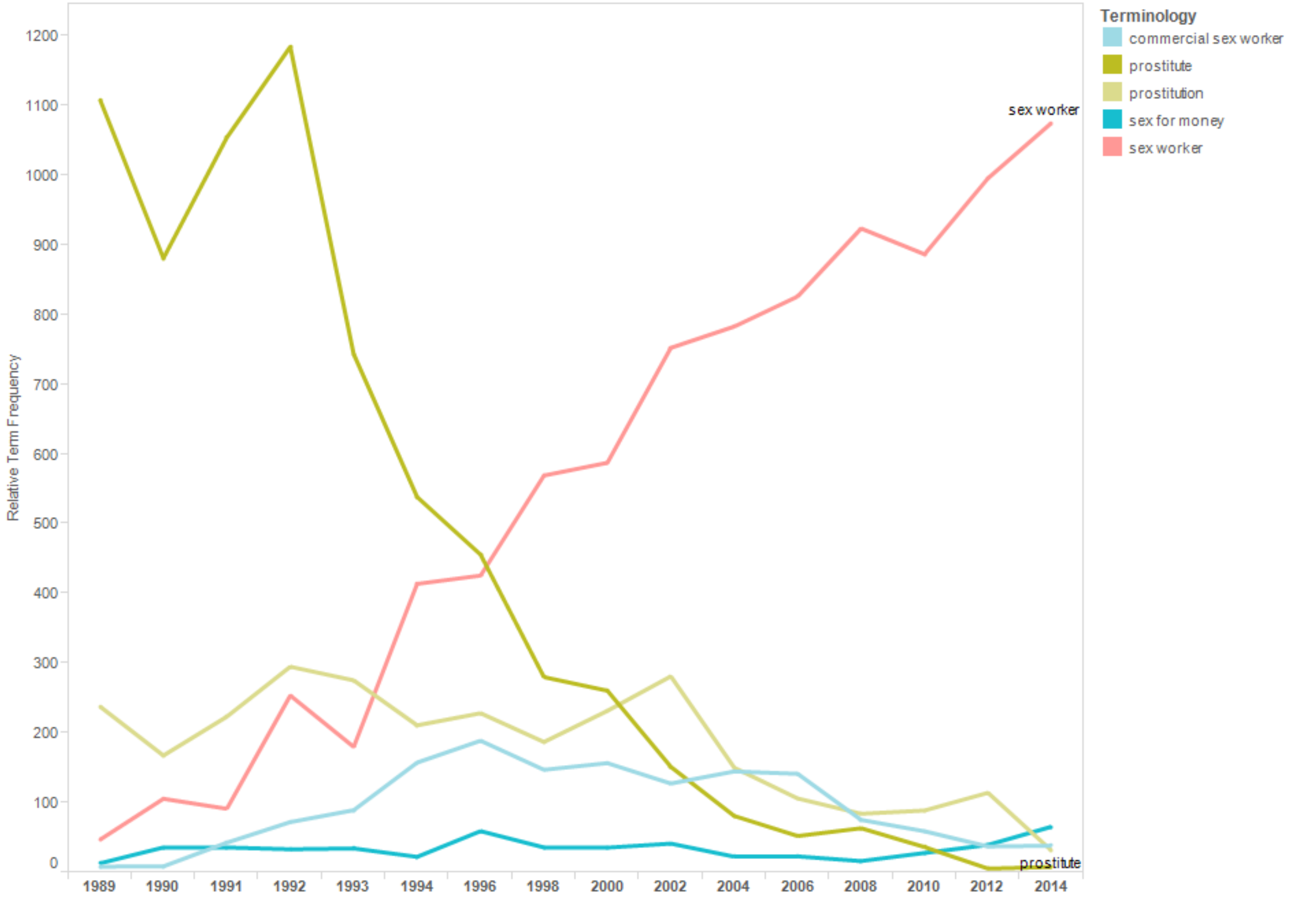


Figure 6. Word cloud for terms associated with the “alcohol and drug use” category in conference year 1989.

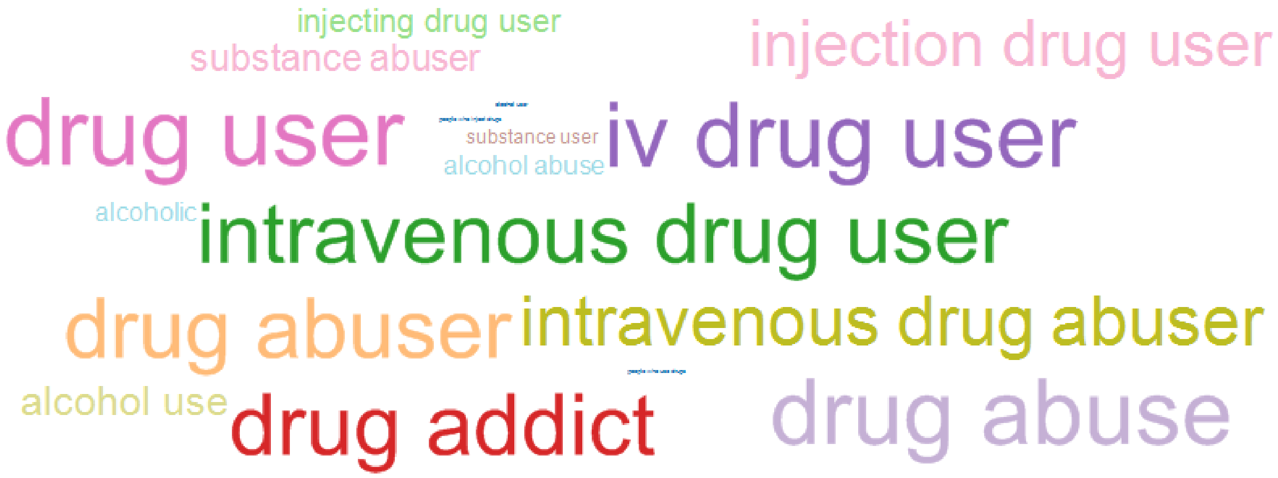


Figure 7. Word cloud for terms associated with the “alcohol and drug use” category in conference year 2014.

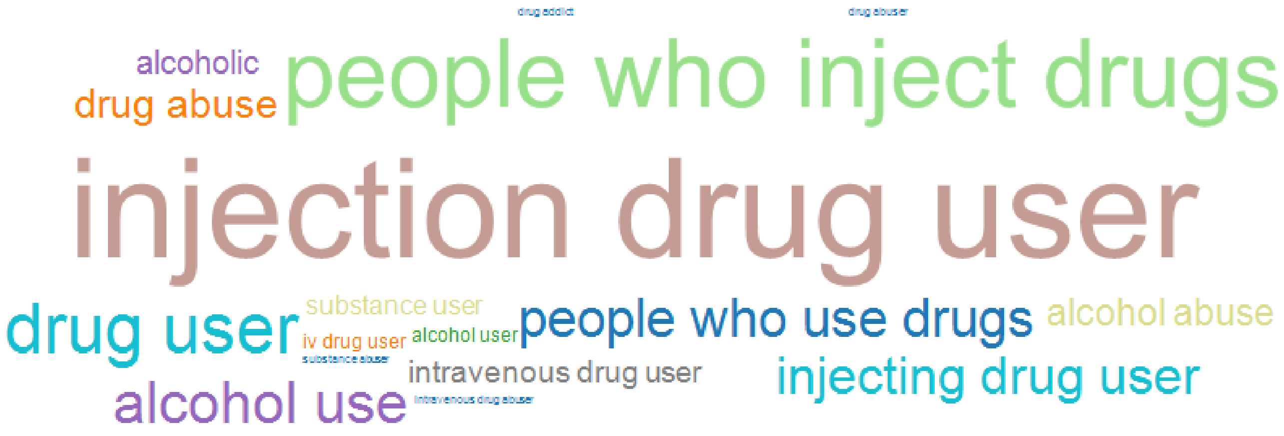


Figure 8. Relative term frequencies for terms associated with the “older adults” category.

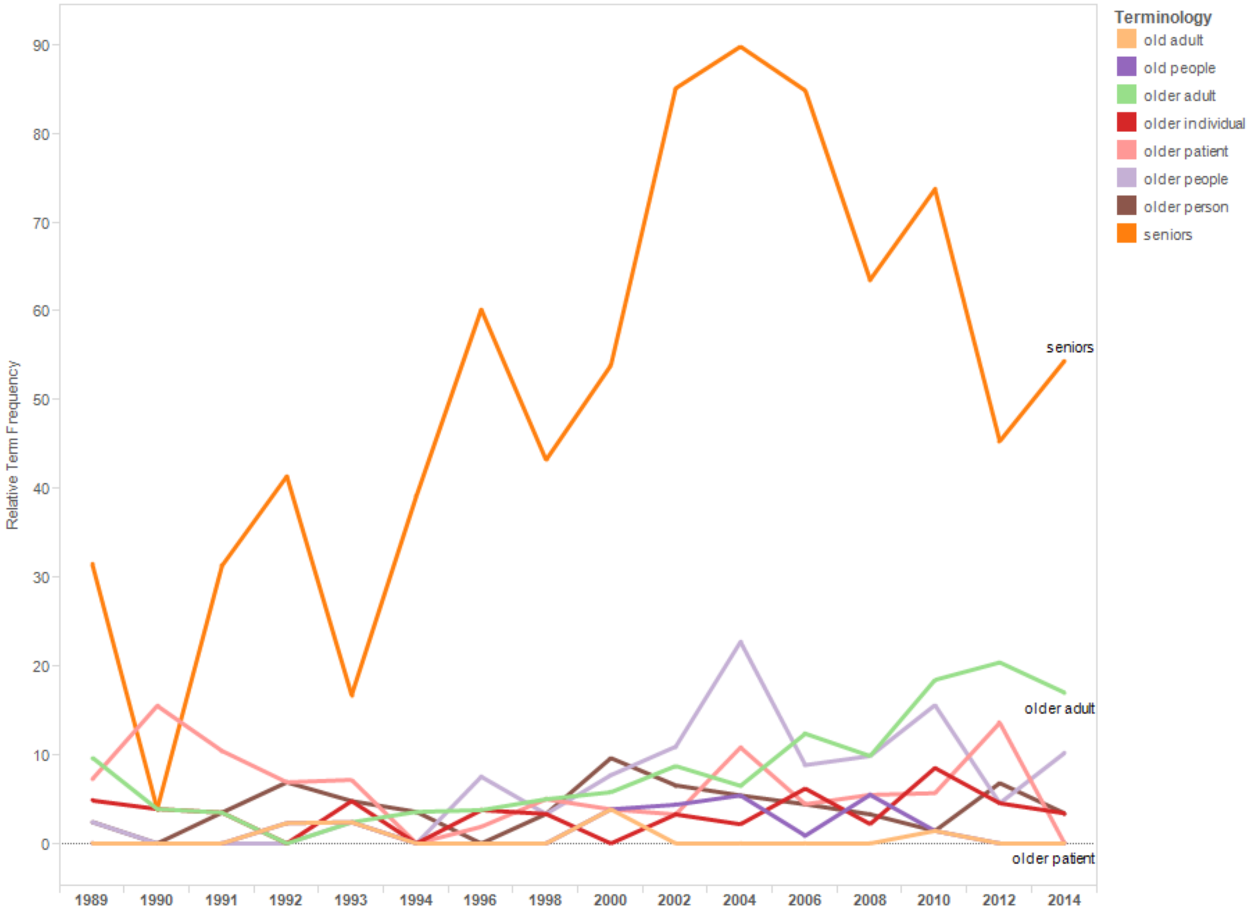


Figure 9. Relative term frequencies for terms associated with “antiretroviral therapy” category.

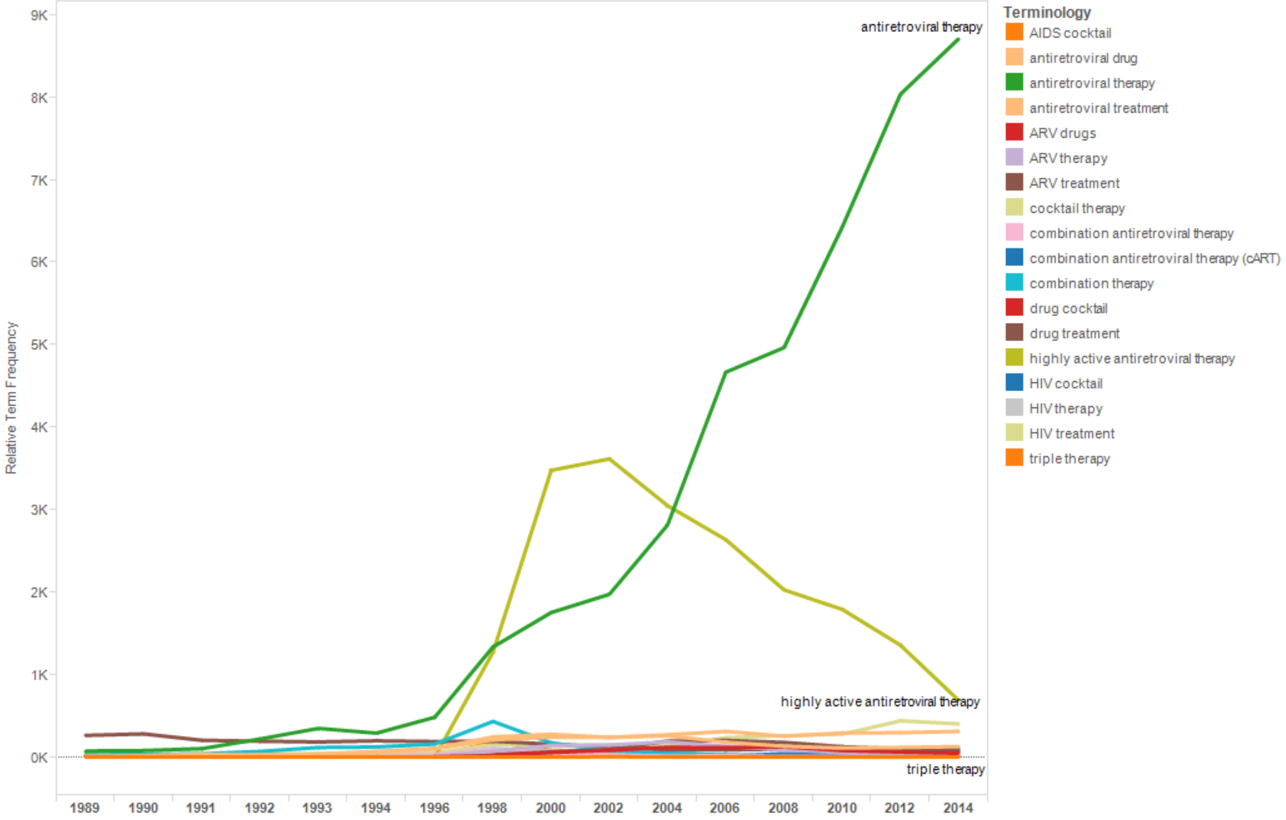
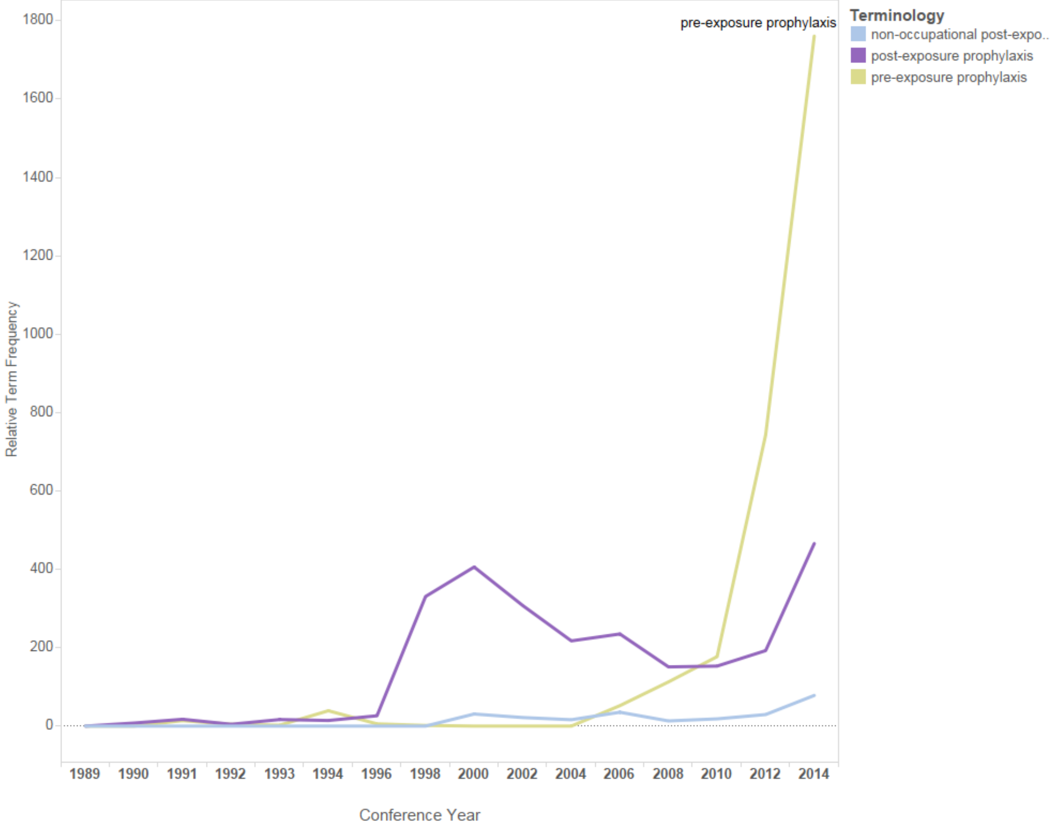


Figure 10. Relative term frequencies for terms associated with the “pre-exposure prophylaxis and post-exposure prophylaxis” category.



Treatment Descriptors

Figure 9 shows a line graph with the relTFs for the terms in the category “antiretroviral therapy.” The relTF of the term “highly active antiretroviral therapy” increased from 1998 to 2002 and then decreased through 2014. The terms “AIDS cocktail” and “HIV cocktail” began appearing in abstracts in the late 1990s, appeared through the mid-2000s, and then were not used again until the late 2000s.

Figure 10 shows a line chart of the relTFs for the terms included in the terminology category “pre-exposure prophylaxis and post-exposure prophylaxis.” Pre-exposure prophylaxis and post-exposure prophylaxis refer to the use of antiretroviral drugs before and after exposure to HIV, respectively, to reduce the risk of HIV infection [15,16]. From 1996 to 2000, the relTF of “post-exposure prophylaxis” increased and then declined from 2000 to 2010. From 2010 to 2014, the relTF of “pre-exposure prophylaxis” and “treatment as prevention” increased.

Inception of Terms

Results of this research also include some terms that appeared in the IAC abstracts only for a limited number of conference years. Use of the term “AIDS epidemic” was dominant from 1989 to 1991 and then declined in use. In contrast, use of the term “HIV epidemic” increased through 2014. Beginning in the mid-1990s, the term “treatment experienced” appeared with increasing frequency in the abstracts.

Use of the term “treatment interruption” began to increase in 1997, appearing in IAC abstracts with increasing frequency until 2004. Similarly, use of the related terms “stopping treatment” and “structured treatment interruption” peaked in 2000 and 2004, respectively, and then precipitously declined in use through 2014. Finally, beginning in the mid-1990s, the terms “continuum of care” and “HIV care continuum” appeared with increasing frequency in the IAC abstracts.

Discussion

Principal Findings

As anticipated, changes in HIV language use across the IAC abstracts illuminated in the Tableau-generated dashboards parallel real-world advances in HIV research and medical practice. Importantly, the changes also highlight the declining stigmatization of both people living with HIV and those at increased risk of HIV exposure and areas where greater sensitivity in terminology would be beneficial.

People Descriptors

Discovery of the virus that causes AIDS in 1981 [1] spurred research to better understand the natural history of HIV. This research led to the development of HIV tests able to detect HIV along the disease continuum before the onset of AIDS. Accordingly, across the succession of IAC conferences, abstract authors increasingly distinguished between “HIV infection” and “AIDS” as indicated by the declining frequency trends lines for terms such as “AIDS epidemic,” “AIDS patient,” and “HIV/AIDS patient” and the increasing the trend lines for the terms “HIV epidemic,” and “people living with HIV.”

Results reflect progress to destigmatize HIV and to avoid stigmatizing people at risk of HIV exposure. For example, the trend lines for “drug addict” and “prostitute” declined, whereas those for “people who use drugs” and “sex worker” increased. Similarly, the use of “homosexual” and “gay” declined in abstracts across conference years, whereas the use of “men who have sex with men” increased. The increasing use of the term “living with HIV” likely reflects efforts to destigmatize HIV, as well as the transition of HIV from a fatal to chronic disease and attention to distinguish HIV infection from AIDS.

At the same time, the results of this research suggest the need for greater efforts to destigmatize HIV. It is notable that while people living with HIV prefer language that puts the person before the disease (ie, person-first language) [2,8], even in 2014, the two most frequently used status-referents in the abstracts are “HIV positive” and “HIV infected.” As emphasized in an annotation of HIV terminology recommendations, the word “infected” can be particularly stigmatizing [8] (although declining in use, “HIV infected” appears frequently in abstracts through 2014). Similarly, across conference years, the term “senior” (traditionally used to describe people aged 65 years or older) was used more often than “older adult.” A glossary recommending socially neutral language noted that objections to the term might increase as more baby boomers reach the age of 65 years [17].

Some generally positive terminological shifts to avoid stigmatizing people affected by HIV also merit further discussion. For example, growing use of “men who have sex with men” to describe a group disproportionately affected by HIV reduces stigma and expands the descriptor to include men who do not identify as “gay.” At the same time, in an article by Kaplan et al, using “men who have sex with men” as an HIV risk category excludes some identities (eg, trans feminine people, particularly those from non-Euro-Atlantic cultures) [18]. This categorization, in turn, may prevent some individuals whose gender and sexual identities do not fit binary notions from seeking HIV-related services and participating in HIV research.

Treatment Descriptors

Results highlighted advances in HIV treatment that transformed HIV from a fatal to a chronic disease. The first antiretroviral drug for the treatment of HIV was approved by the US Food and Drug Administration (FDA) in 1987 [19]. However, the effectiveness of HIV monotherapy was inhibited by the ability of HIV to mutate and become drug resistant. Consequently, researchers began to investigate combinations of antiretroviral drugs to more effectively suppress the virus. The research to find a multidrug alternative to HIV monotherapy is reflected in use of the terms “AIDS cocktail” and “HIV cocktail,” in the IAC abstracts from the late 1990s to the mid-2000s and again in the late 2000s. With the approval of an increasing number of antiretroviral drugs and a landmark clinical trial definitively demonstrating the benefits of triple drug therapy in 1997, the terms “highly active antiretroviral therapy” or “HAART” and “antiretroviral therapy” or “ART” increasingly appeared in IAC abstracts through 2014 [20]. The term “highly active antiretroviral therapy” appeared first in IAC abstracts in 1998, spiking in frequency of use in 2002, and then declined in use through 2014. In contrast, the term "antiretroviral therapy" appeared in IAC abstracts across conference years. As combination therapy progressed from an HIV treatment strategy under investigation to the universal standard of care, emphasis on its potent antiviral activity because superfluous. Accordingly, use of the abbreviated "antiretroviral therapy" surpassed "highly active antiretroviral therapy" by 2006 and continued to increase as the preferred term through 2014.

The increasing availability and use of HIV drugs is also reflected in the introduction of the term "treatment experienced" in IAC abstracts in 1996. Increasing use of the term through 2014 likely reflects the urgency to investigate treatment challenges facing the growing population of HIV-infected individuals on long-term antiretroviral therapy (ART).

The increasing use of the term "continuum of care" in IAC abstracts beginning in the mid-1990s also points to the transition of HIV to a chronic disease. The term reflects increasing focus on the span of HIV care from detection and diagnosis to long-term treatment with ART.

Changes in term use in the IAC abstracts also reflect definitive findings in HIV research. For example, in 2006, a major HIV study comparing continuous with episodic ART was halted when preliminary findings demonstrated that episodic ART increased the risk of HIV progression [21,22]. Not surprisingly, use of the term "structured treatment interruption" declined abruptly in conference year 2006. Another example is the spike in use of the term "pre-exposure prophylaxis" or "PrEP" in 2012, the year that FDA approved use of the HIV drug Truvada for pre-exposure prophylaxis [23].

Terminology Recommendations

Results of this study demonstrate changes in HIV-related language reflective of historical changes in the HIV epidemic. Coupling results of this quantitative study with findings from related qualitative research [6-8], a number of terminology recommendations are provided as follows:

- Use consistent, biomedically accurate language to describe HIV. For example, use "HIV" and not "HIV/AIDS," which equates the disease with its terminal stage.
- When describing people living with HIV, put the person before the disease (eg, use "person with HIV" and not "HIV-infected person").
- Avoid language that stigmatizes people at increased risk of HIV infection. For example, use "drug user" not "drug abuser" and "sex worker" not "prostitute."
- Consider group and personal preferences when communicating with and about people affected by HIV. For example, follow the example of the individual or group when using terms such as "gay" or "men who have sex with men."
- Avoid language that assigns blame. For example, use "the treatment failed" not "the individual failed treatment."

We conclude that adoption of these recommendations will further efforts to destigmatize HIV and, most importantly, facilitate effective communication between health professionals and people affected by HIV.

Research Recommendations

The following are recommendations for future research aimed at evaluating changes in HIV-related terminology over time:

- Conduct a study evaluating results of these analyses on people affected by HIV, including people living with HIV, HIV clinicians, outreach workers, case managers, and HIV researchers to gain insight on real-world experience with HIV terminology, which will further evaluate the impact of HIV terminology on language perceptions and importance
- Import data from future IAC conference abstracts into Tableau Desktop, and program the system to automatically update the dashboards based on the new data.
- Employ a more robust version of Tableau Desktop to overcome the data import limits of the free version of the visualization software, and expand the HIV terminology corpus for analysis.
- Use the methods described for this project to analyze other datasets of HIV-related terms, and compare the results of the analyses.

Further research is also needed to identify optimal strategies to promote adoption of the terminology recommendations provided. Future studies may also investigate how implementation of the terminology recommendations enhances communication about HIV, especially between people affected by HIV and health care providers.

Study Limitations

The HIV terminology corpus developed for analysis was based on unigrams of interest selected by HIV subject matter experts, an acknowledged limitation of this project. This nonautomated selection of key unigrams was dependent on the subject matter experts' knowledge of HIV and their familiarity with trends in HIV terminology use. Automated term extraction is an efficient method to quickly extract terms from a large dataset; however, the method does not account for contextual differences. Although automated term extraction may have produced more terms for study analysis, manually term extraction allowed us to consider the contextual use of terms, which was especially important for terms with more than one meaning.

The study is limited to terminology used in IAC abstracts through conference year 2014. Although the 2016 IAC had taken place at the time of the study, the conference abstracts were not available for analysis.

This study was also restricted by the limited number of unigrams selected to build the HIV terminology corpus for analysis. The initial selection of unigrams was restricted by the data import limitations of the free Tableau visualization software used and the scope of the project. Expanding the choice of unigrams would have greatly extended the application processing and computational time required and resulted in an overwhelming number of terms from which to select a terminology corpus for analysis.

Conclusions

Results of this study highlight changes in HIV terminology use over 25 years, including the addition, disappearance, and changing use of terms that reflect advances in HIV research and medical practice and efforts to use less stigmatizing language. Coupled with findings from related quantitative research, HIV-related terminology recommendations based on results of this study are included. Adoption of these recommendations will further efforts to use less stigmatizing language, improve retention in care, and facilitate effective communication between health care professionals, the media, and people affected by HIV.

Acknowledgments

The authors would like to thank the International AIDS Society for providing the International AIDS Conference abstracts for use as part of this study.

Conflicts of Interest

None declared.

Multimedia Appendix 1

HIV terminology corpus.

[PDF File (Adobe PDF File), 55KB - publichealth_v4i2e50_app1.pdf]

References

1. Centers for Disease Control and Prevention. 2001 Jun 01. First Report of AIDS URL: https://www.cdc.gov/mmwr/preview/mmwrhtml/mm5021a1.htm [WebCite Cache ID 6r6D916jv]
2. Joint United Nations Programme on HIV (UNAIDS). 2015. UNAIDS Terminology Guidelines URL: http://www.unaids.org/sites/default/files/media_asset/2015_terminology_guidelines_en.pdf [WebCite Cache ID 6rCFElqGa]
3. Kelly JF, Saitz R, Wakeman S. Language, substance use disorders, and policy: the need to reach consensus on an “addiction-ary”. Alcohol Treat Q 2016;34(1):116-123. [doi: 10.1080/07347324.2016.1113103]
4. Michel JB, Shen YK, Aiden AP, Veres A, Gray MK, Google Books Team, et al. Quantitative analysis of culture using millions of digitized books. Science 2011 Jan 14;331(6014):176-182 [FREE Full text] [doi: 10.1126/science.1199644] [Medline: 21163965]
5. Funk ME. Our words, our story: a textual analysis of articles published in the Bulletin of the Medical Library Association/Journal of the Medical Library Association from 1961 to 2010. J Med Libr Assoc 2013 Jan;101(1):12-20 [FREE Full text] [doi: 10.3163/1536-5050.101.1.003] [Medline: 23405042]
6. Mupenda B, Duvall S, Maman S, Pettifor A, Holub C, Taylor E, et al. Terms used for people living with HIV in the Democratic Republic of the Congo. Qual Health Res 2014 Feb;24(2):209-216 [FREE Full text] [doi: 10.1177/1049732313519869] [Medline: 24463633]
7. Duby Z, Hartmann M, Mahaka I, Munaiwa O, Nabukeera J, Vilakazi N, et al. Lost in translation: language, terminology, and understanding of penile-anal intercourse in an HIV prevention trial in South Africa, Uganda, and Zimbabwe. J Sex Res 2016;53(9):1096-1106 [FREE Full text] [doi: 10.1080/00224499.2015.1069784] [Medline: 26566583]
8. Dilmitis S, Edwards O, Hull B, Margolese S, Mason N, Namiba A, et al. Language, identity and HIV: why do we keep talking about the responsible and responsive use of language? Language matters. J Int AIDS Soc 2012;15(Suppl 2):17990. [doi: 10.7448/IAS.15.4.17990]
9. International AIDS Society. Conferences URL: https://www.iasociety.org/Conferences [WebCite Cache ID 6w4x9Q6D2]
10. Ramos J. CiteSeerX.: Rutgers University; 2003. Using TF-IDF to determine word relevance in document queries URL: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.121.1424&rep=rep1&type=pdf
11. Nguyen TD, Kan MY. Keyphrase extraction in scientific publications. In: Goh DH, Cao TH, Sølvberg IT, Rasmussen E, editors. Asian Digital Libraries. Looking Back 10 Years and Forging New Frontiers. ICADL 2007. Lecture Notes in Computer Science, vol 4822. Berlin, Heidelberg: Springer; 2007.
12. Mazanetz MP, Marmon RJ, Reisser CB, Morao I. Drug discovery applications for KNIME: an open source data mining platform. Curr Top Med Chem 2012;12(18):1965-1979. [Medline: 23110532]
13. Knime. 2017. URL: https://www.knime.com/ [accessed 2017-12-29] [WebCite Cache ID 6w4xrIQpo]
14. Tableau. URL: https://www.tableau.com/ [accessed 2017-12-29] [WebCite Cache ID 6w4yCTPjb]
15. AIDSinfo. 2017. HIV Prevention: Pre-Exposure Prophylaxis (PrEP) URL: https://aidsinfo.nih.gov/understanding-hiv-aids/fact-sheets/20/85/pre-exposure-prophylaxis--prep- [WebCite Cache ID 6yU7HcMhT]
16. AIDSinfo. 2017. HIV Prevention: Post-Exposure Prophylaxis (PEP) URL: https://aidsinfo.nih.gov/understanding-hiv-aids/fact-sheets/20/87/post-exposure-prophylaxis--pep- [WebCite Cache ID 6yU7IckCI]
17. The Newsletter of the Journalists Exchange on Aging. Age Beat URL: http://web.mnstate.edu/hanson/Writing%20about%20older%20people.pdf [WebCite Cache ID 6yU7KjRLe]

18. Kaplan RL, Sevelius J, Ribeiro K. In the name of brevity: the problem with binary HIV risk categories. Glob Public Health 2016;11(7-8):824-834 [FREE Full text] [doi: 10.1080/17441692.2015.1136346] [Medline: 26824592]
19. HIV.gov. A Timeline of HIV URL: https://www.hiv.gov/hiv-basics/overview/history/hiv-and-aids-timeline [WebCite Cache ID 6rCFeabu2]
20. Gulick RM, Mellors JW, Havlir D, Eron JJ, Gonzalez C, McMahon D, et al. Treatment with indinavir, zidovudine, and lamivudine in adults with human immunodeficiency virus infection and prior antiretroviral therapy. N Engl J Med 1997 Sep 11;337(11):734-739. [doi: 10.1056/NEJM199709113371102] [Medline: 9287228]
21. National Institute of Allergy and Infectious Diseases. Clinical Alert: International HIV Trial Finds Continuous Antiretroviral Therapy Superior to Episodic Therapy URL: https://www.nih.gov/news-events/news-releases/international-hiv/aids-trial-finds-continuous-antiretroviral-therapy-superior-episodic-therapy [WebCite Cache ID 6rCG6lhtI]
22. ClinicalTrials.gov. A Comparison of Two Ways to Manage Anti-HIV Treatment (The SMART Study) URL: https://clinicaltrials.gov/ct2/show/NCT00027352 [WebCite Cache ID 6rDjRKJRP]
23. U.S. Food and Drug Administration. FDA approves first drug for reducing the risk of sexually acquired HIV infection URL: https://www.fda.gov/ForConsumers/ConsumerUpdates/ucm311821.htm [WebCite Cache ID 6rDje8ldS]

Abbreviations

ART: antiretroviral therapy
FDA: Food and Drug Administration
IAC: International AIDS Conference
RelTF: relative term frequency
TF: term frequency
TF-IDF: term frequency-inverse document frequency

Edited by G Eysenbach; submitted 27.07.17; peer-reviewed by K Claborn, A Jadhav, K Doing-Harris; comments to author 23.11.17; revised version received 18.01.18; accepted 14.03.18; published 04.05.18.

Please cite as:
Dancy-Scott N, Dutcher GA, Keselman A, Hochstein C, Copty C, Ben-Senia D, Rajan S, Asencio MG, Choi JJ
Trends in HIV Terminology: Text Mining and Data Visualization Assessment of International AIDS Conference Abstracts Over 25 Years
JMIR Public Health Surveill 2018;4(2):e50
URL: http://publichealth.jmir.org/2018/2/e50/
doi:10.2196/publichealth.8552
PMID:29728344

Original Paper

Development of an Agent-Based Model to Investigate the Impact of HIV Self-Testing Programs on Men Who Have Sex With Men in Atlanta and Seattle

Wei Luo[1*], PhD; David A Katz[2*], MPH, PhD; Deven T Hamilton[3], MPH, PhD; Jennie McKenney[4], PhD; Samuel M Jenness[4], PhD; Steven M Goodreau[5], PhD; Joanne D Stekler[2], MD, MPH; Eli S Rosenberg[6], PhD; Patrick S Sullivan[4], DVM, PhD; Susan Cassels[7*], MPH, PhD

[1]Spatial Analysis Research Center, School of Geographical Sciences and Urban Planning, Arizona State University, Tempe, AZ, United States
[2]Department of Medicine, University of Washington, Seattle, WA, United States
[3]Center for Studies in Demography and Ecology, University of Washington, Seattle, WA, United States
[4]Department of Epidemiology, Rollins School of Public Health, Emory University, Atlanta, GA, United States
[5]Center for Studies in Demography and Ecology, Department of Anthropology, University of Washington, Seattle, WA, United States
[6]School of Public Health, Epidemiology and Biostatistics, Rensselaer, Albany, NY, United States
[7]Department of Geography, University of California Santa Barbara, Santa Barbara, CA, United States
[*]these authors contributed equally

Corresponding Author:
Wei Luo, PhD
Spatial Analysis Research Center
School of Geographical Sciences and Urban Planning
Arizona State University
5th Floor
975 S Myrtle Avenue
Tempe, AZ, 85281
United States
Phone: 1 7164722021
Email: wluo23@asu.edu

Abstract

Background: In the United States HIV epidemic, men who have sex with men (MSM) remain the most profoundly affected group. Prevention science is increasingly being organized around HIV testing as a launch point into an HIV prevention continuum for MSM who are not living with HIV and into an HIV care continuum for MSM who are living with HIV. An increasing HIV testing frequency among MSM might decrease future HIV infections by linking men who are living with HIV to antiretroviral care, resulting in viral suppression. Distributing HIV self-test (HIVST) kits is a strategy aimed at increasing HIV testing. Our previous modeling work suggests that the impact of HIV self-tests on transmission dynamics will depend not only on the frequency of tests and testers' behaviors but also on the epidemiological and testing characteristics of the population.

Objective: The objective of our study was to develop an agent-based model to inform public health strategies for promoting safe and effective HIV self-tests to decrease the HIV incidence among MSM in Atlanta, GA, and Seattle, WA, cities representing profoundly different epidemiological settings.

Methods: We adapted and extended a network- and agent-based stochastic simulation model of HIV transmission dynamics that was developed and parameterized to investigate racial disparities in HIV prevalence among MSM in Atlanta. The extension comprised several activities: adding a new set of model parameters for Seattle MSM; adding new parameters for tester types (ie, regular, risk-based, opportunistic-only, or never testers); adding parameters for simplified pre-exposure prophylaxis uptake following negative results for HIV tests; and developing a conceptual framework for the ways in which the provision of HIV self-tests might change testing behaviors. We derived city-specific parameters from previous cohort and cross-sectional studies on MSM in Atlanta and Seattle. Each simulated population comprised 10,000 MSM and targeted HIV prevalences are equivalent to 28% and 11% in Atlanta and Seattle, respectively.

Results: Previous studies provided sufficient data to estimate the model parameters representing nuanced HIV testing patterns and HIV self-test distribution. We calibrated the models to simulate the epidemics representing Atlanta and Seattle, including matching the expected stable HIV prevalence. The revised model facilitated the estimation of changes in 10-year HIV incidence based on counterfactual scenarios of HIV self-test distribution strategies and their impact on testing behaviors.

Conclusions: We demonstrated that the extension of an existing agent-based HIV transmission model was sufficient to simulate the HIV epidemics among MSM in Atlanta and Seattle, to accommodate a more nuanced depiction of HIV testing behaviors than previous models, and to serve as a platform to investigate how HIV self-tests might impact testing and HIV transmission patterns among MSM in Atlanta and Seattle. In our future studies, we will use the model to test how different HIV self-test distribution strategies might affect HIV incidence among MSM.

(JMIR Public Health Surveill 2018;4(2):e58) doi:10.2196/publichealth.9357

KEYWORDS

HIV; men who have sex with men; pre-exposure prophylaxis

Introduction

To date, HIV testing remains the cornerstone of HIV prevention and care. Approximately 15% of 1.2 million individuals living with HIV in the United States are unaware of their infection [1,2], and some estimates suggest that these individuals account for almost one-third of all sexual transmissions in the United States [3]. HIV-positive individuals can be linked to HIV care and antiretroviral therapy (ART), which can improve individual health outcomes and decrease the likelihood of transmission [4]. In contrast, HIV-negative individuals can be linked to HIV prevention strategies, such as pre-exposure prophylaxis (PrEP) [5]. In the United States, men who have sex with men (MSM) constitute a high-risk group for acquiring HIV, and CDC recommends that MSM should undergo tests for HIV, at least, annually [6]. Despite this recommendation, only 56% of US MSM were reported to have undergone the tests in 2015 [7]. The low rate of testing has hindered linking the requisite number of MSM to treatment and prevention of HIV to bring down the rate of new infections below the effective reproductive number. Consequently, HIV diagnoses among young MSM, especially young black MSM in the United States, has been increasing [8].

In 2012, the United States Food and Drug Administration approved the first HIV self-test (HIVST), the OraQuick In-Home HIV Test (OraQuick; OraSure Technologies, Inc; Bethlehem, PA, USA). HIVSTs are the tests that can be entirely performed by the person undergoing the test for HIV, from specimen collection to reading and interpreting test results. One rationale for OraQuick's approval was that the availability of self-testing—a testing modality that provides convenience, privacy, and anonymity—might decrease the number of people living with HIV who remain unaware of their HIV status [9]. HIVST promises to increase the frequency and coverage of testing in several ways. Self-testing offers a way to reach not only people who are in need of HIV testing but also those who are either not linked to medical care or hesitant to attend testing in community-based settings [10]. In addition, HIVST could provide opportunities to supplement current, clinic-based testing to elevate the testing frequency for MSM. However, HIVST using oral fluids has an extended window period, during which individuals recently infected with HIV might test falsely negative [11], which might incorrectly reassure people with recently acquired HIV infection; furthermore, such individuals are highly infectious [12]. Such people might delay treatment initiation and expose their uninfected partners to HIV. Moreover, they might postpone their next test, further delaying both diagnosis and treatment. Furthermore, people who test positive using HIVST might not be linked to care as frequently as those who test in more traditional settings, such as clinics.

Typically, HIVSTs have been found to be highly acceptable among MSM [13], and studies have established the feasibility and acceptability of various strategies for distributing HIVST kits among MSM [14], including social and sexual networks [15,16], geosocial networking apps [17], bathhouses [18], vending machines [19], vouchers [20], mass distribution at LGBTQ–focused events (lesbian, gay, bisexual, transgender, and queer, LGBTQ) [21], or Web-based strategies [22]. However, the use of HIVSTs in the general MSM population or outside of research studies remains partially investigated. Studies among Australian and Spanish MSM before local regulatory approval and among MSM receiving partner services for newly diagnosed HIV in New York City reported extremely low histories of prior HIVST use (<2.5%) [23-25]. Conversely, surveillance data from King County, WA, suggested that approximately 20% of MSM had used an HIVST by 2015 [21]. Empirical patterns of HIVST use and the extent to which men will supplement or replace clinic-based testing with HIVST will be elucidated in future studies. The objective of developing our present model was to investigate how different patterns of use might affect the HIV incidence and, thus, inform how to promote self-testing.

Our research groups have conducted prior modeling work in several areas related to HIV testing. Katz et al [26] developed a deterministic, continuous-time model of HIV transmission dynamics, suggesting that any replacement of clinic-based tests capable of detecting recent HIV infection with self-testing leads to an increased HIV prevalence among Seattle MSM, irrespective of the impact of self-tests on testing frequency. However, one limitation of this model was that it investigated replacing clinic-based tests only and did not consider other ways in which MSM may use HIVST [27]. Furthermore, there was lack of stochasticity in outcomes, infeasibility of parameterizing certain network structures, and exponential complexity with the linear addition of parameters. It would not be feasible to represent the level of heterogeneity in testing qualities, profiles of testers, and testing scenarios in a compartmental model, nor could the model account for specific sex acts and subsequent seroadaptive behavior based on the combination of test results. Khanna et al [28] reported the potential benefits of individualized HIV testing programs for decreasing HIV transmissions among MSM, and Delaney et al [29] used agent-based models to evaluate the potential value of focusing testing programs on MSM who had never been tested before. Both studies highlighted the potential benefits of HIV testing programs at the individual level to decrease the HIV prevalence among MSM; however, these agent-based models did not explicitly account for the use of HIVSTs or nuanced HIV testing behaviors.

Modeling offers an opportunity to process the available information from studies of self-testing, inform the design of self-testing interventions, compare how self-testing could be most effectively used in different epidemiological settings, and estimate the likely impact of scaling self-testing programs for MSM. Thus, we aimed to develop an agent-based model to investigate the potential impact of these tests on the HIV epidemic among MSM from two different epidemic settings in the United States—Atlanta, GA, and Seattle, WA. Approaches like ours are efficient, cost-effective, and complementary to community-level randomized controlled trials for assessing interventions like self-testing, especially in the setting of broad availability of self-tests and diverse programs, including self-tests being provided by health programs. Briefly, we aimed to develop a network-based mathematical model of HIV transmission dynamics in susceptible populations of MSM in Atlanta and Seattle.

Methods

Study Design

We adapted and extended a previously published, dynamic, stochastic network model of HIV transmission dynamics, which was originally designed to elucidate racial disparities in HIV prevalence among MSM in Atlanta [30]. The original model comprised key aspects of HIV transmission dynamics among MSM, including the occurrence of condomless anal intercourse (CAI) and related HIV transmission probabilities, sexual partnership characteristics (eg, main or casual partnerships), HIV testing frequency, and HIV care seeking. In the original model, MSM possessed fixed (eg, race and circumcision status) and dynamic (eg, age and infection status) attributes. In addition, HIV-infected men possessed additional dynamic attributes (eg, diagnosis status, treatment status, infection stage, and plasma HIV RNA level). Consistent with prior studies [31,32], sexual contact networks used separable temporal exponential random graph models (STERGMs) [33], a flexible statistical framework for simulating partnership formation and dissolution across networks [34], facilitating one to match data on the complex cross-sectional network structure, as well as reported relational durations. We implemented STERGMs in the R package suite *statnet* [35] and *EpiModel* [36]. The specific parameters for the sexual behavior developed by the domain and city are presented in Appendix Table A1.

HIV Transmission and Progression

In the model, sexual contacts occurred on three networks, each of which shared the same set of nodes (people) but represented a different relationship type: main partners, casual partners, and one-time partners. Partnership formation depended on the following four predictors: partnership types (main, casual, or one-time partners), degree distribution (number of ongoing partners for each individual), age homophily (partners with similar age), and sexual role segregation (only those men who engaged in receptive anal sex could pair with men who engaged in exclusively insertive anal sex and vice versa). There was a constant relationship dissolution hazard for main and casual partnerships based on the median duration of each type (Multimedia Appendix 1, Table A1).

HIV progression followed both a natural trajectory of disease, in the absence of ART, and an ART-mediated trajectory [37]. In the absence of ART, HIV viral loads progressed in three stages: peak viremia during acute HIV infection (21 days); a set point viral load during the clinical latency (42 days); and a subsequent increase, resulting in AIDS and disease-induced mortality (728 days) [38]. In serodiscordant pairs, HIV transmission probabilities included those mediated by the viral load of the positive partner [39], condom use [40], the presence of the CCR5-Δ32 genetic allele [41,42], receptive versus insertive sexual position of the HIV-negative partner [31], and the circumcision status of an insertive HIV-negative partner [43]. After infection, men were assigned to three clinical care trajectories: those who never initiate treatment; those who initiate treatment and become partially suppressed; and those who initiate treatment and become entirely suppressed. The three trajectories exhibited different rates of infectiousness, HIV diagnosis, ART initiation, HIV viral suppression, and progression to AIDS and death to match empirical estimates of the prevalence of these states [44]. Furthermore, ART decreased the viral load and its associated transmission risk [4] and extended the life span [45].

HIV Pre-Exposure Prophylaxis

We included PrEP use and adherence as additional features of the model. At any given time, 23.4% of HIV-negative MSM in the Seattle model and 11% of HIV-negative men in the Atlanta model were on PrEP according to Darcy Rao (oral communication, May 2017). If the proportion of HIV-negative MSM on PrEP decreased below the threshold coverage because of discontinuation or seroconversion, newly tested HIV-negative MSM were allowed to start PrEP. In addition, MSM on PrEP received a diagnostic HIV test at regular intervals (3 months). In fact, MSM who started PrEP were assigned a fixed adherence profile that reflected an average weekly dosage using data from the US PrEP Demo Project [46] weighted by race or ethnicity using methods from Jenness et al [32], which investigated the impact of the implementation of the US Centers for Disease Control and Prevention’s PrEP guidelines on the national HIV epidemic among MSM. For the Atlanta model, we assigned 21.1% of men as nonadherent (0 doses), 7.0% taking <2 doses/week, 10.0% taking 2-3 doses/week, and 61.9% taking ≥4 doses/week; for the Seattle model, we assigned 14.4% as nonadherent, 4.1% taking <2 doses/week, 5.3% taking 2-3 doses/week, and 76.2% taking ≥4 doses/week [32]. The probability of HIV acquisition per sex act was decreased according to the level of adherence. Little evidence for risk behavior changes among MSM while on PrEP was found in studies predating the development of this model [46-49]; thus, we focused on the PrEP coverage and adherence and did not model such behavior changes.

HIV Testing: Clinic and Self-Testing

To investigate the possible impacts of HIV self-testing strategies on HIV epidemic dynamics in future models, we added more complex parameters describing clinic-based HIV testing behaviors to the original model to create a new baseline model. As existing patterns of the testing behavior vary among MSM, we conceptualized this baseline model with four tester types [50] (Table 1): (1) those who do not test in any setting ("never testers"); (2) those who test for HIV, but without a regular testing interval and without considering sexual episodes (eg, testing outreach at Pride festivals or when seeking medical care for a non-HIV or STD-related reason; "opportunistic-only testers"); (3) those who test for HIV regularly, irrespective of sexual behaviors ("regular testers"); and (4) those who test for HIV in response to a specific sexual episode (eg, CAI with a partner of unknown HIV serostatus; "risk-based testers"). Both regular and risk-based testers could also have an opportunity to become opportunistic testers and test outside of their regular testing interval or in the absence of a specific sexual episode, respectively, when opportunities were available. In addition, many public health agencies recommend different testing frequencies for MSM at higher versus lower risk for HIV acquisition [51-54], and evidence suggests that on an average, MSM at higher risk test more frequently [55-60]. Consequently, we stratified regular testers into high vs low anal intercourse (AI) frequency groups to allow for differential testing frequencies by risk.

The four categories of HIV tester types (Table 1) in the baseline model provide a framework to investigate the efficacy of different HIV self-testing scenarios for both Atlanta and Seattle. In the next stage of this research, we will implement two different self-testing intervention strategies: replacement testing and supplementary testing (Table 2). In the replacement testing scenarios, we will evaluate the proportion of HIV infections averted when HIVSTs replace clinic-based opportunistic tests, regular tests, and risk-based tests, respectively and simultaneously. In addition, we will assess the impact of replacing 25% and 50% of clinic-based tests with HIVSTs, assuming a 90-day window period (duration between when a person might have been exposed to HIV and when a test can give an accurate result) for self-tests [61].

We will also model the proportion of HIV infections averted when self-tests supplement clinic-based tests. In these scenarios, never testers and opportunistic-only testers take one or two supplementary HIVSTs randomly each year. Regular testers can supplement clinic-based tests with HIVSTs in two ways.

Table 1. HIV testing typology for the baseline model (clinic test only) of HIV transmission dynamics among men who have sex with men in Seattle and Atlanta.

Tester type	Baseline testing behavior	Opportunistic testing behavior
Never testers	None	None
Opportunistic-only testers	None	Likelihood of testing when presented with an opportunity to test (varies by the tester type)
Regular testers	Test interval varies by HIV risk category (defined by high vs low AI[a] frequency)	
Risk-based testers	Testing likelihood and time to test varies by three types of events: (1) CAI[b] in non - main partnership, (2) CAI within known serodiscordant partnership, and (3) Acquisition of new main partner	

[a] AI: anal intercourse.

[b] CAI: condomless anal intercourse.

Table 2. Scenarios for modeling uptake of HIV self-testing among men who have sex with men: replacement versus supplementary testing (to be compared with the baseline model).

Tester type	Replacement scenarios	Supplementation scenarios
Never testers	N/A[a]	• 1 test per year • 2 tests per year
Opportunistic-only testers	• 25% replacement • 50% replacement	• 1 additional test per year • 2 additional tests per year
Risk-based testers	• 25% replacement • 50% replacement	• 10% additional probability of testing after risk event • 20% additional probability
Regular testers	• 25% replacement • 50% replacement	• 1 additional test per year at random time • Decrease the intertest interval by adding 1 test per year

[a] N/A: not applicable.

First, they can supplement clinic tests with one additional HIVST randomly during a year. Second, they can test more frequently (ie, shorter intertest interval) using both clinic tests and self-tests. For example, if a regular tester tests three times a year in a clinic, his intertest interval would be 120 days; if he supplements this with one self-test per year, his intertest interval would be 90 days, and he would have a 25% chance that each test would be a self-test. Risk-based testers supplement their clinic-based tests by increasing the probability of risk-based tests by 10% and 20%. For example, if a risk tester has a 40% chance of taking a clinic-based test after a risk event and self-testing increases the probability of risk-based testing by 10%, his likelihood of testing becomes 50% with a 40% chance of taking a clinic-based test and a 10% chance of self-testing.

Results

Data Parameterization

We hypothesized that the impact of HIVST on MSM will differ in cities with different underlying patterns of HIV testing, access to HIV prevention and care, and epidemiological characteristics of MSM. King County (home to Seattle) is one of the first jurisdictions to reach the World Health Organization’s 90-90-90 goals [62] and has witnessed a rapid scale-up of PrEP [63]. HIV prevalence among MSM in Seattle is estimated to be 11% according to Dr Susan Buskin (personal communication, October 2017), and the rate of new HIV diagnoses has been declining over the last decade [64]. In contrast, the HIV prevalence among MSM in Atlanta is estimated to be much higher at 28% [65], with fewer MSM aware of their HIV infection status [66], exhibiting less access to PrEP and a more diverse population with significant disparities in HIV by race [67]. The data sources for parameterizing our model for each city are presented in Table 3, with additional detail in the Appendix Table A1. As the original model was structured to assess racial differences in the incidence, where race-stratified estimates were available, we used those estimates to create race-weighted composite estimates for the overall population in each city. Both HIV testing and network-based behavioral data were collected in a prospective cohort study [65] and an egocentric sexual network study [68] of MSM in Atlanta and cross-sectional surveys of MSM in Seattle [55,69].

We estimated parameters to characterize HIV testing behaviors in both Atlanta and Seattle, including the proportions of the four tester types (Table 4) and mean test intervals and test likelihood for opportunistic-only, regular, and risk-based testers (Table 5). Of note, these HIV testing parameters only apply to MSM who are not on PrEP (MSM on PrEP received a diagnostic HIV test at regular 3-month intervals). For Seattle, the proportions of the four tester types, the mean intervals between regular tests stratified by the frequency of AI, and the likelihood that opportunistic testers will seize a testing opportunity were estimated using data obtained from 361 HIV-negative and unknown-status MSM aged 18-39 years who participated in the Seattle Pride Survey [63] in 2013 and 2014 and reported living in Washington State. The Pride Survey is an annual convenience sample of self-identified MSM enrolled from along the Seattle Pride Parade route to complete a self-administered survey to monitor HIV risk behaviors and the uptake of treatment and prevention interventions. The 2013 and 2014 surveys included the following question: “How many HIV tests did you have in the last two years?” To calculate the mean interval in days between regular tests, we divided 2 years (730.5 days) by the mean number of HIV tests reported in the last 2 years among the participants reporting their last test as a “routine test,” stratified by the number of AI partners in the last year (0-2 vs 3+ partners). In addition, we calculated the likelihood that opportunistic-only testers will seize a testing opportunity by dividing the mean number of HIV tests reported in the last 2 years by the expected number of opportunities to test in a 2-year period (4 opportunities: 2 each year, based on expert opinion) among participants who reported the reason for their last test as something other than risk-based or routine.

For MSM in Atlanta, the proportions of the four tester types were estimated using data from the Man Project [68] and InvolveMENt [65]. The Man Project was a cross-sectional, chain-referral sexual network study among 314 MSM; however, InvolveMENt was a prospective cohort study among 803 black and white non-Hispanic MSM aged 18-39 years. Participants for both the studies completed self-administered computer-based questionnaires that assessed demographics as well as risk and prevention behaviors for both themselves and their most recent sex partners (up to the last 10 partners in the last 12 months for the Man Project and the last 5 partners in the previous 6 months for InvolveMENt). We evaluated the mean interval in days between regular tests using data from the PUMA Survey [70], an internet-based survey of MSM in the United States conducted from November 30, 2010 to December 19, 2010 (*n*=1251). Men who reported regularly testing for HIV were asked how often they tested (in months). Then, we calculated the mean number of months between regular tests, but unlike for the parameters of MSM in Seattle, we were unable to stratify this interval by the number of AI partners. Furthermore, we calculated the likelihood that opportunistic-only testers seized an opportunity from data collected in the Annual American Men’s Internet Survey of Behaviors of Men who have Sex with Men in the United States (AMIS) [7,69,71]. AMIS is an annual cross-sectional behavioral survey of MSM in the United States in which participants complete a Web-based survey that includes questions in the following domains: demographics, sexual behavior, HIV testing history, drug and alcohol use, and HIV prevention services exposure. As with the Seattle data, the likelihood was calculated among participants who self-reported living in Georgia and reported testing for reasons other than risk or routine by dividing the mean number of HIV tests reported among them in the last 2 years by the expected number of opportunities in that period.

Where local data for parameterization were unavailable for both cities, we used estimates from national Web-based surveys conducted by research teams involving members of our group. The likelihood of seizing a testing opportunity among regular and risk-based testers was estimated by calculating the proportion of individuals who reported currently testing on a regular schedule and reported that their last test was in response to an opportunity using data from the Internet Ethics and Incentives Study [7,37,69]. Specifically, to parameterize this model, we added questions to the 2016 cycle of AMIS [7,37,69] to ascertain the likelihood of testing following a risk event and the time from the risk event to test. Respondents were asked whether they tested in response to each of the following events: the last time they had CAI with a nonmain partner, the last time they had CAI with a serodiscordant partner, and the last time they started a new relationship. In addition, we calculated the likelihood of testing following each event as the proportion of individuals who reported testing in response to such an event among those who reported the event. Subsequently, men who reported that they last tested in response to CAI with a nonmain partner, CAI with a serodiscordant partner, or starting a new relationship were asked how many weeks following the event they tested. The time from event to test was calculated as the mean number of weeks for each type of event separately.

Model Calibration

Using the above parameters, we simulated the sexual, vital, and epidemiological dynamics of HIV transmission among 10,000 MSM for 50 years using Seattle- and Atlanta-specific parameterizations and, at equilibrium, reproduced the epidemiological and demographic outcomes observed in Seattle and Atlanta, respectively. In addition, we calibrated the baseline models by varying the rate of AI within partnerships to closely match the estimates of HIV prevalence that were 28% in Atlanta and 11% in Seattle among MSM (Figure 1). We used a multiplier of 1.5 in Atlanta and 3.45 in Seattle to adjust the AI rate, which resulted in simulated epidemics equilibrating at 27.24% (95% CI: 27.79-28.35) in Atlanta and 11.55% (95% CI: 11.36-11.73) in Seattle. To select the multiplier, we varied the AI multiplier by 0.01 within preselected ranges of 1.45-1.55 and 3.4-3.5 for Atlanta and Seattle, respectively. Then, we identified the value that resulted in an average equilibrium HIV prevalence (from 100 simulations for each value for 50 years each) closest to the target estimates.

The next steps in our research will be to introduce each self-test scenario into the simulated populations and project epidemic outcomes over a 10-year time horizon. We will further investigate how different test sensitivities and detection windows affect the impact of HIVST.

Table 3. Data sources for parameterization of a model of HIV transmission dynamics among MSM in Atlanta, GA, and Seattle, WA.

Parameters	Atlanta Data Sources	Seattle Data Sources
Testing and treatment	InvolveMENt [65,72-74]	Medical Monitoring Project [75,76], HIV Surveillance [62,77]
Tester type	AMIS[a] [7,69,71], Man Project [68], InvolveMENt [65]	Seattle Pride Survey [63]
Mean interval between tests among regular testers	PUMA[b] Survey [70]	Seattle Pride Survey [63]
Likelihood of seizing the testing opportunity for risk-based, regular, and opportunity-only testers	Internet Ethics and Incentives Study [78], American Men's Internet Survey [7,69,71]	Internet Ethics and Incentives Study [78], Seattle Pride Survey [63]
Risk-based testing: the likelihood of testing after event and time from event to test	American Men's Internet Survey [7,69,71]	American Men's Internet Survey [7,69,71]
Sexual behavior (ie, versatility, condom use, disclosure); sexual network attributes	Man Project [68], InvolveMENt [65]	Mobile Study (S. Cassels, Personal Communication, November 2017)
Prevalence of circumcision	InvolveMENt [65]	National HIV Behavioral Surveillance [66,79,80]
PrEP coverage	PUMA Survey [70]	Washington HIV Prevention Project
Expected coital frequency within partnerships	PUMA Survey [70]	Washington HIV Prevention Project
Racial or ethnic distribution of MSM for weighting parameters estimates	Goodreau et al [30]	King County Population Estimates [81]
CCR5-Δ32 prevalence	Marmor et al [82]	Zimmerman et al [41] and Marmor et al [82]

[a] AMIS: Annual American Men’s Internet Survey of Behaviors of Men who have Sex with Men in the United States.

[b] PUMA: Prevention Umbrella for MSM in the Americas.

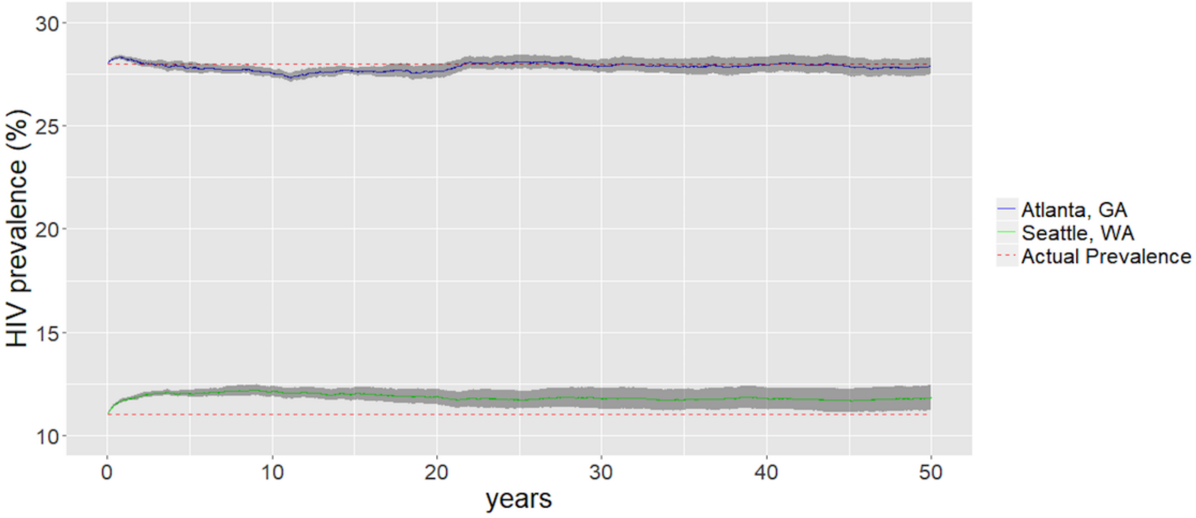
Table 4. Estimated proportions of four tester types in Atlanta and Seattle.

Tester type	Atlanta	Seattle
Never testers	3.5%	2.5%
Opportunistic-only testers	37.0%	13.8%
Regular testers	44.0%	64.9%
Risk-based testers	15.5%	18.8%

Table 5. Mean test intervals and test likelihood for opportunistic-only, regular, and risk-based testers in both Atlanta and Seattle.

Tester Type	Atlanta	Seattle
Opportunistic-only testers		
Interval between opportunities to take an opportunistic test	183 days	183 days
Likelihood of seizing testing opportunity: Opportunistic-only testers	0.629	0.764
Regular testers		
Interval between regular tests for high AI frequency group	224 days	151 days
Interval between regular tests for low AI frequency group	224 days	372 days
Likelihood of seizing testing opportunity: Regular tester	0.095	0.095
Risk-based testers		
Likelihood of testing after event: CAI in non-main partnership	0.339	0.359
Likelihood of testing after event: CAI within known serodiscordant partnership	0.520	0.538
Likelihood of testing after event: Acquisition of new main partner	0.349	0.375
Time from event to test: CAI in non-main partnership	39.2 days	39.2 days
Time from event to test: CAI within known serodiscordant partnership	43.4 days	43.4 days
Time from event to test: Acquisition of new main partner	56.7 days	56.7 days
Likelihood of seizing testing opportunity: Risk-based tester	0.095	0.095

Figure 1. Simulated HIV prevalence plots to produce baseline epidemics for Atlanta, GA, and Seattle, WA, with 95% CIs (gray band). The actual HIV prevalence among the men who have sex with men population was 28% in Atlanta and 11% in Seattle.



Discussion

This study adapted our previous dynamic, network-based epidemic model of HIV transmission dynamics [30] and conceptualized and implemented a baseline model as a platform for investigating different HIVST intervention strategies at the individual level. These dynamic network models are ideal for investigating the effects of complex, interacting interventions along structured and evolving contact networks [31]. Our primary improvement was adding four different types of testers to this model (never testers, opportunistic-only testers, regular testers, and risk-based testers) and deriving estimates of prevalence as well as patterns of testing behaviors for each tester type. It will enable us to better reflect the range of testing behaviors among MSM than other models, which usually assume that all men test in the same manner, and will enable us to assess the potential impact of targeting self-testing to specific types of testers. In addition, we estimated the parameters related to HIV behaviors among MSM in Atlanta and Seattle to compare the impact of different epidemic settings on HIVST intervention strategies. Our enhanced model will be used to examine self-testing intervention strategies to determine the most effective ways to promote HIVST for MSM in the United States. Furthermore, our models offer great promise to account for novel testing strategies and to inform public health approaches for the promotion of safe and effective HIV self-testing strategies in two divergent settings, Seattle and Atlanta.

As with all models, ours has several limitations. First, to match the observed HIV prevalence in Seattle and Atlanta, our models required calibration for one behavioral parameter, the frequency of AI within partnerships. Not all factors that contribute to the differential HIV transmission dynamics could be included in this model; thus, the variation in the frequency of AI within partnerships could also represent the variation in the processes not included in the model. For example, we adopted a simplistic approach to PrEP use in our model. PrEP indications, uptake, and use over time are complex behaviors, and individuals who initiate PrEP might be nonrandomly associated with certain tester types. However, we modeled PrEP use as a consistent coverage fraction and did not have detailed data to parameterize how PrEP use interacted with our newly parameterized testing categories. Our models were based on data collected from numerous data sources, rather than one comprehensive, representative source for each city, and in some cases, local data sources were unavailable; thus, we had to rely on national surveys that are unlikely to be representative of either city. Second, testing parameters relied primarily on self-reported behaviors. Thus, the findings from the model will be considered as hypothetical scenarios to illustrate how similar approaches to HIV testing, self-testing in particular, can affect HIV transmission dynamics in various ways in different epidemic settings. Furthermore, our model should not be used as a prediction tool to assess the trajectory of the epidemic for specific subpopulations. Finally, the testing categories are static in the model. Hence, further investigations are warranted to elucidate how individual testing behavior varies over the life span.

HIVST provides a new opportunity to reach out to people who are in need of HIV testing. Understanding how to maximize the potential of HIV self-testing—and limit its risks—in the current landscape of high-sensitivity and high-throughput HIV testing is important. However, evaluations of self-testing programs have been limited in that it is difficult to assess who might be using the kits, what the results of the tests are, and, thus, what impact the self-tests might have on the epidemic. We executed a new modeling project to focus on the potential roles of self-tests in increasing HIV testing and decreasing the HIV incidence among MSM and to develop a model that would facilitate the assessment of how local epidemic characteristics might affect the impact of self-testing programs. The objectives were attained by developing parallel models for Seattle and Atlanta, which are different in their epidemics among MSM regarding testing behaviors; access to HIV testing, prevention, and care; and demographic and epidemic characteristics [83]. The finalization of this protocol will enable us to examine the population-level effects of HIVST; test different HIVST intervention strategies, such as replacement and supplementary testing, and mass and targeted distributions; and determine populations and settings where HIVST might have the greatest impact [84]. Overall, this study represents one more step forward in unfolding the process of improving the realism of epidemic models, and parameterizing them using rich local data, to best determine the optimal strategies for promoting population health in different communities.

Acknowledgments

This material is supported by funding from the NIH/NICHD R21 HD080523. We acknowledge Ms Darcy Rao and Dr Sara Glick from the University of Washington, Dr Julia Hood and Dr Susan Buskin from Public Health–Seattle & King County, and the Washington State Department of Health for providing data to parameterize the Seattle model. We also acknowledge Dr Travis Sanchez and Ms Maria Zlotorzynska from Emory University for providing data from the Annual American Men's Internet Survey for model parameterization. In addition, we acknowledge funding from the studies that developed the model that we extended (R21-HD075662) and that provided the data: National Institute of Mental Health (R01-MH085600), Eunice Kennedy Shriver National Institute of Child Health & Human Development (R01-HD067111, R21HD074502), Emory CFAR (P30 AI050409), and the MAC AIDS Fund. Finally, we thank the University of Washington Center for Studies in Demography and Ecology (R24HD042828) for providing computational resources.

Conflicts of Interest

None declared.

Multimedia Appendix 1

The specific parameters for sexual behavior developed by domain and city.

[XLSX File (Microsoft Excel File), 23KB - publichealth_v4i2e58_app1.xlsx]

References

1. Hall HI, An Q, Tang T, Song R, Chen M, Green T, et al. Prevalence of Diagnosed and Undiagnosed HIV Infection--United States, 2008-2012. MMWR Morb Mortal Wkly Rep 2015 Jun 26;64(24):657-662 [FREE Full text] [Medline: 26110835]
2. Centers FDCP. Monitoring selected national HIV prevention and care objectives by using HIV surveillance data-United States and 6 dependent areas-2012. Atlanta, GA: US Department of Health and Human Services, CDC; 2014. URL: https://www.cdc.gov/hiv/pdf/surveillance_report_vol_19_no_3.pdf [accessed 2018-05-23] [WebCite Cache ID 6zgBdGimY]
3. Skarbinski J, Rosenberg E, Paz-Bailey G, Hall HI, Rose CE, Viall AH, et al. Human immunodeficiency virus transmission at each step of the care continuum in the United States. JAMA Intern Med 2015 Apr;175(4):588-596. [doi: 10.1001/jamainternmed.2014.8180] [Medline: 25706928]
4. Cohen MS, Chen YQ, McCauley M, Gamble T, Hosseinipour MC, Kumarasamy N, et al. Prevention of HIV-1 infection with early antiretroviral therapy. N Engl J Med 2011 Aug 11;365(6):493-505 [FREE Full text] [doi: 10.1056/NEJMoa1105243] [Medline: 21767103]
5. Cohen MS, Chen YQ, McCauley M, Gamble T, Hosseinipour MC, Kumarasamy N, et al. Prevention of HIV-1 infection with early antiretroviral therapy. N Engl J Med 2011 Aug 11;365(6):493-505 [FREE Full text] [doi: 10.1056/NEJMoa1105243] [Medline: 21767103]
6. Centers for Disease Control and Prevention, 2017. Recommendations for HIV Screening of Gay, Bisexual, and Other Men Who Have Sex with Men – United States. 2017. Morbidity and Mortality Weekly Reports URL: https://www.cdc.gov/mmwr/volumes/66/wr/mm6631a3.htm [accessed 2018-06-01] [WebCite Cache ID 6zrAEAYkg]
7. Zlotorzynska M, Sullivan P, Sanchez T. The Annual American Men's Internet Survey of Behaviors of Men Who Have Sex With Men in the United States: 2015 Key Indicators Report. JMIR Public Health Surveill 2017 Mar 25;3(1):e13 [FREE Full text] [doi: 10.2196/publichealth.7119] [Medline: 28356240]
8. Rosenberg ES, Grey JA, Sanchez TH, Sullivan PS. Rates of Prevalent HIV Infection, Prevalent Diagnoses, and New Diagnoses Among Men Who Have Sex With Men in US States, Metropolitan Statistical Areas, and Counties, 2012-2013. JMIR Public Health Surveill 2016 May 17;2(1):e22 [FREE Full text] [doi: 10.2196/publichealth.5684] [Medline: 27244769]
9. U.S. Food and Drug Administration. Available: Accessed 15 January. 2018. OraQuick® In-Home HIV Testummary of Safety and Effectiveness URL: https://www.fda.gov/downloads/BiologicsBloodVaccines/BloodBloodProducts/ApprovedProducts/PremarketApprovalsPMAs/UCM312534.pdf [accessed 2018-06-01] [WebCite Cache ID 6zrATMG82]
10. Walensky RP, Paltiel AD. Rapid HIV testing at home: does it solve a problem or create one? Ann Intern Med 2006 Sep 19;145(6):459-462. [Medline: 16983134]
11. Stekler JD, O'Neal JD, Lane A, Swanson F, Maenza J, Stevens CE, et al. Relative accuracy of serum, whole blood, and oral fluid HIV tests among Seattle men who have sex with men. J Clin Virol 2013 Dec;58 Suppl 1:e119-e122 [FREE Full text] [doi: 10.1016/j.jcv.2013.09.018] [Medline: 24342471]

12. Wawer MJ, Gray RH, Sewankambo NK, Serwadda D, Li X, Laeyendecker O, et al. Rates of HIV-1 transmission per coital act, by stage of HIV-1 infection, in Rakai, Uganda. J Infect Dis 2005 May 01;191(9):1403-1409. [doi: 10.1086/429411] [Medline: 15809897]
13. Figueroa C, Johnson C, Verster A, Baggaley R. Attitudes and acceptability on HIV self-testing among key populations: a literature review. AIDS Behav 2015 Nov;19(11):1949-1965 [FREE Full text] [doi: 10.1007/s10461-015-1097-8] [Medline: 26054390]
14. Estem KS, Catania J, Klausner JD. HIV self-testing: a review of current implementation and fidelity. Curr HIV/AIDS Rep 2016 Apr;13(2):107-115. [doi: 10.1007/s11904-016-0307-y] [Medline: 26879653]
15. Carballo-Diéguez A, Frasca T, Balan I, Ibitoye M, Dolezal C. Use of a rapid HIV home test prevents HIV exposure in a high risk sample of men who have sex with men. AIDS Behav 2012 Oct;16(7):1753-1760 [FREE Full text] [doi: 10.1007/s10461-012-0274-2] [Medline: 22893194]
16. Lippman SA, Lane T, Rabede O, Gilmore H, Chen Y, Mlotshwa N, et al. High acceptability and increased HIV-testing frequency after introduction of HIV self-testing and network distribution among South African MSM. J Acquir Immune Defic Syndr 2018 Mar 01;77(3):279-287. [doi: 10.1097/QAI.0000000000001601] [Medline: 29210826]
17. Huang E, Marlin RW, Young SD, Medline A, Klausner JD. Using Grindr, a smartphone social-networking application, to increase HIV self-testing among black and Latino men who have sex with men in Los Angeles, 2014. AIDS Educ Prev 2016 Aug;28(4):341-350. [doi: 10.1521/aeap.2016.28.4.341] [Medline: 27427928]
18. Woods WJ, Lippman SA, Agnew E, Carroll S, Binson D. Bathhouse distribution of HIV self-testing kits reaches diverse, high-risk population. AIDS Care 2016;28 Suppl 1:111-113 [FREE Full text] [doi: 10.1080/09540121.2016.1146399] [Medline: 26883730]
19. Young SD, Klausner J, Fynn R, Bolan R. Electronic vending machines for dispensing rapid HIV self-testing kits: a case study. AIDS Care 2014 Feb;26(2):267-269 [FREE Full text] [doi: 10.1080/09540121.2013.808732] [Medline: 23777528]
20. Marlin RW, Young SD, Bristow CC, Wilson G, Rodriguez J, Ortiz J, et al. Piloting an HIV self-test kit voucher program to raise serostatus awareness of high-risk African Americans, Los Angeles. BMC Public Health 2014 Nov 26;14:1226 [FREE Full text] [doi: 10.1186/1471-2458-14-1226] [Medline: 25427749]
21. Katz D, Bennett A, Dombrowski J, Hood J, Buskin S, Golden M. 2015 National HIV Prevention Conference, Atlanta, GA. Monitoring the Population-Level Impact of HIV Self-Testing through HIV SurveillancePartner Services URL: http://hivst.org/evidence/monitoring-the-population-level-impact-of-hiv-self-testing-through-hiv-surveillance-and-partner-services [accessed 2018-06-01] [WebCite Cache ID 6zrB3ptIn]
22. Edelstein Z, Salcuni P, Tsoi B. CROI 2017; Seattle, WA. FeasibilityReach of a HIV Self-Test (HIVST) Giveaway, New York City, 2015-16 URL: https://www1.nyc.gov/assets/doh/downloads/pdf/dires/hiv-self-test-giveaway.pdf [accessed 2018-06-01] [WebCite Cache ID 6zrAluP9r]
23. Koutentakis K, Rosales-Statkus ME, Hoyos J, Fernández-Balbuena S, Ruiz M, Agustí C, Madrid HIV self-testing group. Knowledge and use of unauthorized HIV self-test kits among men who have sex with men in Spain, following approval of an over-the-counter self-test in the U.S: a cross-sectional study. BMC Public Health 2016 Dec 08;16:532 [FREE Full text] [doi: 10.1186/s12889-016-3204-9] [Medline: 27390840]
24. Prestage G, Zablotska I, Bavinton B, Grulich A, Keen P, Murphy D, et al. Previous and future use of HIV self-testing: a survey of Australian gay and bisexual men. Sex Health 2016 Feb;13(1):55-62. [doi: 10.1071/SH15099] [Medline: 26502289]
25. Udeagu CN, Shah S, Molochevski M. Men who have sex with men seek timely human immunodeficiency virus confirmation and care after rapid human immunodeficiency virus self-test: data from partner services program, New York City. Sex Transm Dis 2017 Oct;44(10):608-612. [doi: 10.1097/OLQ.0000000000000648] [Medline: 28876324]
26. Katz DA, Cassels SL, Stekler JD. Replacing clinic-based tests with home-use tests may increase HIV prevalence among Seattle men who have sex with men: evidence from a mathematical model. Sex Transm Dis 2014 Jan;41(1):2-9 [FREE Full text] [doi: 10.1097/OLQ.0000000000000046] [Medline: 24335742]
27. Jamil MS, Prestage G, Fairley CK, Grulich AE, Smith KS, Chen M, et al. Effect of availability of HIV self-testing on HIV testing frequency in gay and bisexual men at high risk of infection (FORTH): a waiting-list randomised controlled trial. Lancet HIV 2017 Jun;4(6):e241-e250. [doi: 10.1016/S2352-3018(17)30023-1] [Medline: 28219619]
28. Khanna A, Goodreau SM, Wohlfeiler D, Daar E, Little S, Gorbach PM. Individualized diagnosis interventions can add significant effectiveness in reducing human immunodeficiency virus incidence among men who have sex with men: insights from Southern California. Ann Epidemiol 2015 Jan;25(1):1-6 [FREE Full text] [doi: 10.1016/j.annepidem.2014.09.012] [Medline: 25453725]
29. Delaney KP, Rosenberg ES, Kramer MR, Waller LA, Sullivan PS. Optimizing human immunodeficiency virus testing interventions for men who have sex with men in the United States: a modeling study. Open Forum Infect Dis 2015 Dec;2(4):ofv153 [FREE Full text] [doi: 10.1093/ofid/ofv153] [Medline: 26613096]
30. Goodreau SM, Rosenberg ES, Jenness SM, Luisi N, Stansfield SE, Millett GA, et al. Sources of racial disparities in HIV prevalence in men who have sex with men in Atlanta, GA, USA: a modelling study. Lancet HIV 2017 Jul;4(7):e311-e320. [doi: 10.1016/S2352-3018(17)30067-X] [Medline: 28431923]

31. Goodreau SM, Carnegie NB, Vittinghoff E, Lama JR, Sanchez J, Grinsztejn B, et al. What drives the US and Peruvian HIV epidemics in men who have sex with men (MSM)? PLoS One 2012;7(11):e50522 [FREE Full text] [doi: 10.1371/journal.pone.0050522] [Medline: 23209768]
32. Jenness SM, Goodreau SM, Rosenberg E, Beylerian EN, Hoover KW, Smith DK, et al. Impact of the Centers for Disease Control's HIV Preexposure Prophylaxis Guidelines for men who have sex with men in the United States. J Infect Dis 2016 Dec 15;214(12):1800-1807. [doi: 10.1093/infdis/jiw223] [Medline: 27418048]
33. Krivitsky PN, Morris M. Inference for social network models from egocentrically sampled data, with application to understanding persistent racial disparities in HIV prevalence in the US. Ann. Appl. Stat 2017 Mar;11(1):427-455. [doi: 10.1214/16-AOAS1010]
34. Krivitsky PN, Handcock MS. A separable model for dynamic networks. J R Stat Soc Series B Stat Methodol 2014 Jan 01;76(1):29-46 [FREE Full text] [doi: 10.1111/rssb.12014] [Medline: 24443639]
35. Handcock MS, Hunter DR, Butts CT, Goodreau SM, Morris M. statnet: Software Tools for the Representation, Visualization, Analysis and Simulation of Network Data. J Stat Softw 2008;24(1):1548-7660 [FREE Full text] [Medline: 18618019]
36. Jenness SM, Goodreau SM, Morris M. EpiModel: An R Package for Mathematical Modeling of Infectious Disease over Networks. J Stat Softw 2018 Apr;84 [FREE Full text] [doi: 10.18637/jss.v084.i08] [Medline: 29731699]
37. Carnegie NB, Goodreau SM, Liu A, Vittinghoff E, Sanchez J, Lama JR, et al. Targeting pre-exposure prophylaxis among men who have sex with men in the United States and Peru: partnership types, contact rates, and sexual role. J Acquir Immune Defic Syndr 2015 May 01;69(1):119-125 [FREE Full text] [doi: 10.1097/QAI.0000000000000555] [Medline: 25942463]
38. Pilcher CD, Price MA, Hoffman IF, Galvin S, Martinson FEA, Kazembe PN, et al. Frequent detection of acute primary HIV infection in men in Malawi. AIDS 2004 Feb 20;18(3):517-524. [Medline: 15090805]
39. Hughes JP, Baeten JM, Lingappa JR, Magaret AS, Wald A, de BG, Partners in Prevention HSV/HIV Transmission Study Team. Determinants of per-coital-act HIV-1 infectivity among African HIV-1-serodiscordant couples. J Infect Dis 2012 Feb 01;205(3):358-365 [FREE Full text] [doi: 10.1093/infdis/jir747] [Medline: 22241800]
40. Weller S, Davis K. Condom effectiveness in reducing heterosexual HIV transmission. Cochrane Database Syst Rev 2002(1):CD003255. [doi: 10.1002/14651858.CD003255] [Medline: 11869658]
41. Zimmerman PA, Buckler-White A, Alkhatib G, Spalding T, Kubofcik J, Combadiere C, et al. Inherited resistance to HIV-1 conferred by an inactivating mutation in CC chemokine receptor 5: studies in populations with contrasting clinical phenotypes, defined racial background, and quantified risk. Mol Med 1997 Jan;3(1):23-36 [FREE Full text] [Medline: 9132277]
42. Marmor M, Sheppard HW, Donnell D, Bozeman S, Celum C, Buchbinder S, HIV Network for Prevention Trials Vaccine Preparedness Protocol Team. Homozygous and heterozygous CCR5-Delta32 genotypes are associated with resistance to HIV infection. J Acquir Immune Defic Syndr 2001 Aug 15;27(5):472-481. [Medline: 11511825]
43. Wiysonge CS, Kongnyuy EJ, Shey M, Muula AS, Navti OB, Akl EA, et al. Male circumcision for prevention of homosexual acquisition of HIV in men. Cochrane Database Syst Rev 2011 Jun 15(6):CD007496. [doi: 10.1002/14651858.CD007496.pub2] [Medline: 21678366]
44. Sullivan PS, Rosenberg ES, Sanchez TH, Kelley CF, Luisi N, Cooper HL, et al. Explaining racial disparities in HIV incidence in black and white men who have sex with men in Atlanta, GA: a prospective observational cohort study. Ann Epidemiol 2015 Jun;25(6):445-454. [doi: 10.1016/j.annepidem.2015.03.006] [Medline: 25911980]
45. Goodreau SM, Carnegie NB, Vittinghoff E, Lama JR, Fuchs JD, Sanchez J, et al. Can male circumcision have an impact on the HIV epidemic in men who have sex with men? PLoS One 2014;9(7):e102960 [FREE Full text] [doi: 10.1371/journal.pone.0102960] [Medline: 25076493]
46. Liu AY, Cohen SE, Vittinghoff E, Anderson PL, Doblecki-Lewis S, Bacon O, et al. Preexposure prophylaxis for HIV infection integrated with municipal- and community-based sexual health services. JAMA Intern Med 2016 Jan;176(1):75-84 [FREE Full text] [doi: 10.1001/jamainternmed.2015.4683] [Medline: 26571482]
47. Grant RM, Anderson PL, McMahan V, Liu A, Amico KR, Mehrotra M, et al. Uptake of pre-exposure prophylaxis, sexual practices, and HIV incidence in men and transgender women who have sex with men: a cohort study. Lancet Infect Dis 2014 Sep;14(9):820-829. [doi: 10.1016/S1473-3099(14)70847-3] [Medline: 25065857]
48. McCormack S, Dunn DT, Desai M, Dolling DI, Gafos M, Gilson R, et al. Pre-exposure prophylaxis to prevent the acquisition of HIV-1 infection (PROUD): effectiveness results from the pilot phase of a pragmatic open-label randomised trial. Lancet 2016 Jan 2;387(10013):53-60 [FREE Full text] [doi: 10.1016/S0140-6736(15)00056-2] [Medline: 26364263]
49. Marcus JL, Glidden DV, Mayer KH, Liu AY, Buchbinder SP, Amico KR, et al. No evidence of sexual risk compensation in the iPrEx trial of daily oral HIV preexposure prophylaxis. PLoS One 2013;8(12):e81997 [FREE Full text] [doi: 10.1371/journal.pone.0081997] [Medline: 24367497]
50. Hussen SA, Stephenson R, del RC, Wilton L, Wallace J, Wheeler D, HPTN 061 Protocol Team. HIV testing patterns among black men who have sex with men: a qualitative typology. PLoS One 2013;8(9):e75382 [FREE Full text] [doi: 10.1371/journal.pone.0075382] [Medline: 24069408]
51. British HA, British AOSH. UK National Guidelines for HIV Testing. 2008. British Infection Society URL: http://www.bhiva.org/documents/Guidelines/Testing/GlinesHIVTest08.pdf [accessed 2018-06-01] [WebCite Cache ID 6zrAthitB]

52. Sexually Transmissible Infections in Gay Men Action Group. Australian Sexually Transmitted Infection & HIV Testing Guidelines. 2014. URL: https://stipu.nsw.gov.au/wp-content/uploads/STIGMA_Testing_Guidelines_Final_v5-2.pdf [WebCite Cache ID 6zrBA5JpA]
53. Public HS&KC. Available at: Accessed 11 Aug. 2017. HIV testing and STD screening recommendations for men who have sex with men (MSM) URL: https://www.kingcounty.gov/depts/health/communicable-diseases/hiv-std/providers/screening-testing-msm.aspx [accessed 2018-06-01] [WebCite Cache ID 6zrBCThtl]
54. Workowski KA, Berman SM. Centers for Disease Control and Prevention sexually transmitted diseases treatment guidelines. Clin Infect Dis 2007 Apr 01;44 Suppl 3:S73-S76. [doi: 10.1086/511430] [Medline: 17342670]
55. Katz DA, Dombrowski JC, Swanson F, Buskin SE, Golden MR, Stekler JD. HIV intertest interval among MSM in King County, Washington. Sex Transm Infect 2013 Feb;89(1):32-37 [FREE Full text] [doi: 10.1136/sextrans-2011-050470] [Medline: 22563016]
56. MacKellar DA, Valleroy LA, Anderson JE, Behel S, Secura GM, Bingham T, et al. Recent HIV testing among young men who have sex with men: correlates, contexts, and HIV seroconversion. Sex Transm Dis 2006 Mar;33(3):183-192. [doi: 10.1097/01.olq.0000204507.21902.b3] [Medline: 16508526]
57. Manning SE, Thorpe LE, Ramaswamy C, Hajat A, Marx MA, Karpati AM, et al. Estimation of HIV prevalence, risk factors, and testing frequency among sexually active men who have sex with men, aged 18-64 years--New York City, 2002. J Urban Health 2007 Mar;84(2):212-225 [FREE Full text] [doi: 10.1007/s11524-006-9135-5] [Medline: 17295058]
58. Visser M, Heijne JCM, Hogewoning AA, van AF. Frequency and determinants of consistent STI/HIV testing among men who have sex with men testing at STI outpatient clinics in the Netherlands: a longitudinal study. Sex Transm Infect 2017 Sep;93(6):396-403 [FREE Full text] [doi: 10.1136/sextrans-2016-052918] [Medline: 28159917]
59. Johnson BA, McKenney J, Ricca AV, Rosenberg ES, Liu C, Sharma A, et al. Risk factors associated with repeated HIV testing among internet-using men who have sex with men. AIDS Educ Prev 2016 Dec;28(6):511-523. [doi: 10.1521/aeap.2016.28.6.511] [Medline: 27925484]
60. Hoenigl M, Anderson CM, Green N, Mehta SR, Smith DM, Little SJ. Repeat HIV-testing is associated with an increase in behavioral risk among men who have sex with men: a cohort study. BMC Med 2015 Sep 11;13:218 [FREE Full text] [doi: 10.1186/s12916-015-0458-5] [Medline: 26444673]
61. Stekler JD, Ure G, O'Neal JD, Lane A, Swanson F, Maenza J, et al. Performance of Determine Combo and other point-of-care HIV tests among Seattle MSM. J Clin Virol 2016 Mar;76:8-13 [FREE Full text] [doi: 10.1016/j.jcv.2015.12.011] [Medline: 26774543]
62. HIV/ AIDS Epidemiology Unit, Public Health - Seattle & King Countythe Infectious Disease Assessment Unit, Washington State Department of Health. HIV/AIDS Epidemiology Report 2016, Volume. URL: https://www.kingcounty.gov/depts/health/communicable-diseases/hiv-std/patients/epidemiology/~/media/depts/health/communicable-diseases/documents/hivstd/2008-second-half-hiv-aids-epidemiology-annual-report.ashx [WebCite Cache ID 6zrBP3byD]
63. Hood JE, Buskin SE, Dombrowski JC, Kern DA, Barash EA, Katzi DA, et al. Dramatic increase in preexposure prophylaxis use among MSM in Washington state. AIDS 2016 Jan 28;30(3):515-519. [doi: 10.1097/QAD.0000000000000937] [Medline: 26562845]
64. Golden MR, Bennett AB, Dombrowski JC, Buskin SE. Achieving the goals of the National HIV/AIDS strategy: declining HIV diagnoses, improving clinical outcomes, and diminishing racial/ethnic disparities in King County, WA (2004-2013). Sex Transm Dis 2016 May;43(5):269-276. [doi: 10.1097/OLQ.0000000000000434] [Medline: 27100761]
65. Sullivan PS, Peterson J, Rosenberg ES, Kelley CF, Cooper H, Vaughan A, et al. Understanding racial HIV/STI disparities in black and white men who have sex with men: a multilevel approach. PLoS One 2014;9(3):e90514 [FREE Full text] [doi: 10.1371/journal.pone.0090514] [Medline: 24608176]
66. Centers FDCAP. HIV Surveillance Special Report 15. HIV Infection Risk, Prevention, and Testing Behaviors among Men Who Have Sex With Men-National HIV Behavioral Surveillance, 20 U.S. Cities, 2014 URL: https://www.cdc.gov/hiv/pdf/library/reports/surveillance/cdc-hiv-hssr-nhbs-msm-2014.pdf [accessed 2018-06-01] [WebCite Cache ID 6zrBjHpyq]
67. Giustini D. OA Librarian. Open Source, Open Access - And Open Search? URL: https://aidsvu.org/about-aidsvu/ [accessed 2017-11-03] [WebCite Cache ID 6uhaIjixG]
68. Hernández-Romieu AC, Sullivan PS, Rothenberg R, Grey J, Luisi N, Kelley CF, et al. Heterogeneity of HIV prevalence among the sexual networks of black and white men who have sex with men in Atlanta: illuminating a mechanism for increased HIV risk for young black men who have sex with men. Sex Transm Dis 2015 Sep;42(9):505-512 [FREE Full text] [doi: 10.1097/OLQ.0000000000000332] [Medline: 26267877]
69. Sanchez T, Zlotorzynska M, Sineath C, Kahle E, Sullivan P. The Annual American Men's Internet Survey of Behaviors of Men Who have Sex with Men in the United States: 2014 key indicators report. JMIR Public Health Surveill 2016;2(1):e23 [FREE Full text] [doi: 10.2196/publichealth.5476] [Medline: 27244770]
70. Volk JE, Liu A, Vittinghoff E, Irvin R, Kroboth E, Krakower D, et al. Sexual frequency and planning among at-risk men who have sex with men in the United States: implications for event-based intermittent pre-exposure prophylaxis. J Acquir Immune Defic Syndr 2012 Sep 01;61(1):112-115 [FREE Full text] [doi: 10.1097/QAI.0b013e31825bd87d] [Medline: 22592590]

71. Sanchez TH, Sineath RC, Kahle EM, Tregear SJ, Sullivan PS. The Annual American Men's Internet Survey of Behaviors of Men Who Have Sex With Men in the United States: protocol and key indicators report 2013. JMIR Public Health Surveill 2015;1(1):e3 [FREE Full text] [doi: 10.2196/publichealth.4314] [Medline: 27227126]
72. Rosenberg ES, Millett GA, Sullivan PS, Del RC, Curran JW. Understanding the HIV disparities between black and white men who have sex with men in the USA using the HIV care continuum: a modeling study. Lancet HIV 2014 Dec;1(3):e112-e118 [FREE Full text] [doi: 10.1016/S2352-3018(14)00011-3] [Medline: 25530987]
73. Fiebig EW, Wright DJ, Rawal BD, Garrett PE, Schumacher RT, Peddada L, et al. Dynamics of HIV viremia and antibody seroconversion in plasma donors: implications for diagnosis and staging of primary HIV infection. AIDS 2003 Sep 05;17(13):1871-1879. [doi: 10.1097/01.aids.0000076308.76477.b8] [Medline: 12960819]
74. Beer L, Oster AM, Mattson CL, Skarbinski J, Medical Monitoring Project. Disparities in HIV transmission risk among HIV-infected black and white men who have sex with men, United States, 2009. AIDS 2014 Jan 02;28(1):105-114 [FREE Full text] [doi: 10.1097/QAD.0000000000000021] [Medline: 23942058]
75. Blair JM, Fagan JL, Frazier EL, Do A, Bradley H, Valverde EE, National Center for HIV/AIDS, Viral Hepatitis, STD,TB Prevention, CDC. Behavioral and clinical characteristics of persons receiving medical care for HIV infection - Medical Monitoring Project, United States, 2009. MMWR Suppl 2014 Jun 20;63(5):1-22. [Medline: 24941443]
76. Centers for Disease Control and Prevention. Medical Monitoring Project. 2015. URL: https://www.cdc.gov/hiv/statistics/systems/mmp/index.html [accessed 2018-06-01] [WebCite Cache ID 6zrC3bldx]
77. Centers FDCAP. HIV Surveillance Report. 2015. URL: https://www.cdc.gov/hiv/library/reports/hiv-surveillance.html [accessed 2018-06-01] [WebCite Cache ID 6zrC85yox]
78. Hall EW, Sanchez TH, Stein AD, Stephenson R, Zlotorzynska M, Sineath RC, et al. Use of videos improves informed consent comprehension in web-based surveys among internet-using men who have sex with men: a randomized controlled trial. J Med Internet Res 2017 Mar 06;19(3):e64 [FREE Full text] [doi: 10.2196/jmir.6710] [Medline: 28264794]
79. Centers FDCAP. HIV Surveillance Special Report 8. 2014 Sep. HIV Risk, Prevention, and Testing Behaviors?National HIV Behavioral Surveillance System: Men Who Have Sex With Men, 20 U.S. Cities, 2011 URL: https://www.cdc.gov/hiv/library/reports/surveillance/ [accessed 2018-06-01] [WebCite Cache ID 6zrCHvx8p]
80. Finlayson TJ, Le B, Smith A, Bowles K, Cribbin M, Miles I, Centers for Disease ControlPrevention (CDC). HIV risk, prevention, and testing behaviors among men who have sex with men--National HIV Behavioral Surveillance System, 21 U.S. cities, United States, 2008. MMWR Surveill Summ 2011 Oct 28;60(14):1-34 [FREE Full text] [Medline: 22031280]
81. Washington SOOFM. Accessed 26 May. Estimates of April 1 population by age, sex, race and Hispanic origin URL: https://cdnjs.com/ [accessed 2018-06-03] [WebCite Cache ID 6zuiVmKJx]
82. Marmor M, Sheppard HW, Donnell D, Bozeman S, Celum C, Buchbinder S, HIV Network for Prevention Trials Vaccine Preparedness Protocol Team. Homozygous and heterozygous CCR5-Delta32 genotypes are associated with resistance to HIV infection. J Acquir Immune Defic Syndr 2001 Aug 15;27(5):472-481. [Medline: 11511825]
83. Centers for Disease Control and Prevention. HIV Surveillance Special Report 15. HIV Infection Risk, Prevention, and Testing Behaviors among Men Who Have Sex With Men?National HIV Behavioral Surveillance, 20 U.S. Cities, 2014 URL: https://www.cdc.gov/hiv/pdf/library/reports/surveillance/cdc-hiv-hssr-nhbs-msm-2014.pdf [accessed 2018-06-01] [WebCite Cache ID 6zrCUjfYc]
84. Katz D, Golden M, Stekler J. 2012 National Summit on HIV and Viral Hepatitis Diagnosis. Prevention, and Access to Care. 2012. Point-of-sex testing: intentions of men who have sex with men to use home-use HIV tests with sex partners URL: http://www.hivforum.org/projects/69-general/754-2012-summit-breakout-sessions [accessed 2018-06-01] [WebCite Cache ID 6zrCcsnxa]

Abbreviations

AI: anal intercourse
ART: antiretroviral therapy
CAI: condomless anal intercourse
CDC: Centers for Disease Control and Prevention
HIVST: HIV self-test
MSM: men who have sex with men
PrEP: pre-exposure prophylaxis
PUMA: Prevention Umbrella for MSM in the Americas

Edited by G Eysenbach; submitted 13.11.17; peer-reviewed by V Cambiano, J Salisbury, T Torres; comments to author 17.12.17; revised version received 05.04.18; accepted 24.04.18; published 29.06.18.

Please cite as:
Luo W, Katz DA, Hamilton DT, McKenney J, Jenness SM, Goodreau SM, Stekler JD, Rosenberg ES, Sullivan PS, Cassels S
Development of an Agent-Based Model to Investigate the Impact of HIV Self-Testing Programs on Men Who Have Sex With Men in Atlanta and Seattle
JMIR Public Health Surveill 2018;4(2):e58
URL: http://publichealth.jmir.org/2018/2/e58/
doi:10.2196/publichealth.9357
PMID:29959112

Original Paper

The Acceptability and Feasibility of Implementing a Bio-Behavioral Enhanced Surveillance Tool for Sexually Transmitted Infections in England: Mixed-Methods Study

Sonali Wayal[1,2,3], PhD; David Reid[2,3,4], MSc; Paula B Blomquist[2,3], MSc; Peter Weatherburn[3,4], MSc; Catherine H Mercer[1,3*], PhD; Gwenda Hughes[1,2,3*], PhD

[1]Centre for Population Research in Sexual Health and HIV, Institute for Global Health, University College London, London, United Kingdom

[2]Centre for Infectious Disease Surveillance and Control, HIV & STI Department, Public Health England, London, United Kingdom

[3]The National Institute for Health Research, Health Protection Research Unit in Blood Borne and Sexually Transmitted Infections, University College London / Public Health England / London School of Hygiene & Tropical Medicine, London, United Kingdom

[4]Sigma Research, Public Health, Environments and Society, London School of Hygiene & Tropical Medicine, London, United Kingdom

[*]these authors contributed equally

Corresponding Author:
Sonali Wayal, PhD
Centre for Population Research in Sexual Health and HIV
Institute for Global Health
University College London
3rd Floor
Mortimer Market Centre
London, WC1E6JB
United Kingdom
Phone: 44 02031082064
Email: s.wayal@ucl.ac.uk

Abstract

Background: Sexually transmitted infection (STI) surveillance is vital for tracking the scale and pattern of epidemics; however, it often lacks data on the underlying drivers of STIs.

Objective: This study aimed to assess the acceptability and feasibility of implementing a bio-behavioral enhanced surveillance tool, comprising a self-administered Web-based survey among sexual health clinic attendees, as well as linking this to their electronic health records (EHR) held in England's national STI surveillance system.

Methods: Staff from 19 purposively selected sexual health clinics across England and men who have sex with men and black Caribbeans, because of high STI burden among these groups, were interviewed to assess the acceptability of the proposed bio-behavioral enhanced surveillance tool. Subsequently, sexual health clinic staff invited all attendees to complete a Web-based survey on drivers of STI risk using a study tablet or participants' own digital device. They recorded the number of attendees invited and participants' clinic numbers, which were used to link survey data to the EHR. Participants' online consent was obtained, separately for survey participation and linkage. In postimplementation phase, sexual health clinic staff were reinterviewed to assess the feasibility of implementing the bio-behavioral enhanced surveillance tool. Acceptability and feasibility of implementing the bio-behavioral enhanced surveillance tool were assessed by analyzing these qualitative and quantitative data.

Results: Prior to implementation of the bio-behavioral enhanced surveillance tool, sexual health clinic staff and attendees emphasized the importance of free internet/Wi-Fi access, confidentiality, and anonymity for increasing the acceptability of the bio-behavioral enhanced surveillance tool among attendees. Implementation of the bio-behavioral enhanced surveillance tool across sexual health clinics varied considerably and was influenced by sexual health clinics' culture of prioritization of research and innovation and availability of resources for implementing the surveys. Of the 7367 attendees invited, 85.28% (6283) agreed to participate. Of these, 72.97% (4585/6283) consented to participate in the survey, and 70.62% (4437/6283) were eligible and completed it. Of these, 91.19% (4046/4437) consented to EHR linkage, which did not differ by age or gender but was higher among gay/bisexual men than heterosexual men (95.50%, 722/756 vs 88.31%, 1073/1215; $P<.003$) and lower among black Caribbeans than white participants (87.25%, 568/651 vs 93.89%, 2181/2323; $P<.002$). Linkage was achieved for 88.88% (3596/4046) of consenting participants.

Conclusions: Implementing a bio-behavioral enhanced surveillance tool in sexual health clinics was feasible and acceptable to staff and groups at STI risk; however, ensuring participants’ confidentiality and anonymity and availability of resources is vital. Bio-behavioral enhanced surveillance tools could enable timely collection of detailed behavioral data for effective commissioning of sexual health services.

(JMIR Public Health Surveill 2018;4(2):e52) doi:10.2196/publichealth.9010

KEYWORDS

public health surveillance; sexually transmitted diseases; feasibility studies; electronic health records; Web-based survey

Introduction

Sexually Transmitted Infections’ Surveillance in England

Globally, the burden of sexually transmitted infections (STI) continues to be high [1,2]. STI surveillance is a valuable public health tool to monitor the scale and trends of infections and the effectiveness of prevention strategies. In England, Public Health England (PHE), an executive agency of the Department of Health, manages a mandatory, national STI surveillance system known as GUMCAD, which contains pseudonymized, patient-level, electronic health records (EHR) of STI diagnoses and sexual health services accessed by all sexual health clinic (SHC) attendees, along with their sociodemographic characteristics [3]. PHE is also responsible for providing guidance on the management of STI outbreaks and epidemics [4]. Recent GUMCAD data have shown that between 2007 and 2016, the number of new STI diagnoses in England has increased [5]. Men who have sex with men (MSM), young people, and people of black ethnic minorities, particularly black Caribbeans (BC), bear a disproportionate STI burden and thus are priority groups for STI prevention efforts. Although existing GUMCAD data provide an excellent overview of STI epidemics and variations in subgroups, its interpretation is hampered by the lack of systematically collected information on STI risk behaviors. Enhancements to GUMCAD by collecting data on behavioral indicators of STI risk are planned [6], but these will not (and are not intended to) provide sufficient detail to investigate risk practices and contextual factors associated with neither specific nor evolving epidemics.

Collecting Biological and Behavioral Data for Surveillance of Sexually Transmitted Infections

In the context of HIV, globally, since the late 1990s, tailoring surveillance to the epidemic state of a country and collecting and comparing behavioral and prevalence data have helped better understand the course of epidemics [7]. Similarly, in England, an investigation of sexually transmitted *Shigella flexneri* outbreak in MSM between 2012 and 2013 using face-to-face semistructured quantitative interviews uncovered unexpected risk behaviors and their drivers, and highlighted the value of collecting enhanced behavioral data alongside biological data for informing STI control strategies in response to emergent public health concerns [8]. Thus, a bio-behavioral enhanced surveillance tool (BBEST) could be designed to collect detailed behavioral, attitudinal, and contextual data, depending on the nature of the public health concern, that is, data that will not be available in GUMCAD. However, the acceptability and feasibility of using a BBEST to collect sensitive behavioral data, especially if these relate to illicit behaviors, for example, recreational drug use, from SHC attendees and linking these to EHRs for STI surveillance are unknown. Moreover, peoples’ ability and willingness to accurately respond to sensitive behavioral questions could influence the reliability and validity of such data. Compared with face-to-face and with pen-and-paper interviews, computer-assisted self-interview (CASI) methods have been shown to result in greater disclosure of sensitive behaviors [9]. CASI also reduces item nonresponse, in part due to programmed routing of questions [10], thereby increasing the validity of resulting parameters [11].

Study Aims and Objectives

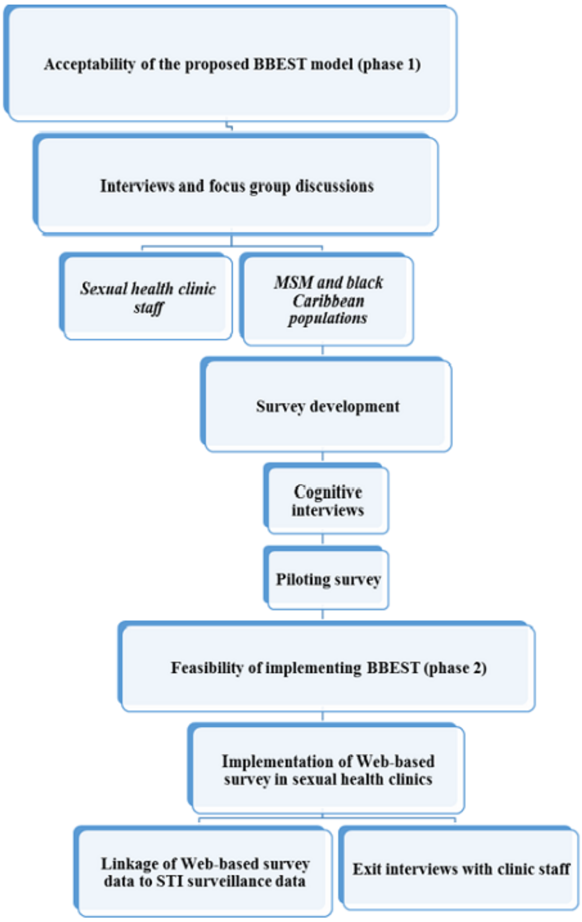
Our aim was to assess the acceptability and feasibility of implementing a BBEST, comprising offering a self-administered Web-based survey to SHC attendees in England, to be completed using digital devices, and subsequently linking their survey responses to their EHR. Although we undertook this assessment in SHCs, our primary focus was on people of BC ethnicity and MSM as exemplar populations because of high STI burden among these groups, as mentioned previously.

Methods

Study Overview

We used a mixed methods study design (Figure 1) comprising phase 1 to assess acceptability of the proposed model of BBEST and phase 2 to examine the feasibility of implementing it. A Community Advisory Group (CAG) and a Steering Group comprising experts and stakeholders in the field of sexual health were set up to guide the study process. The CAG was involved in the development of study materials, including developing study posters, participant information sheet (PIS), and the terminology used in the survey. National Research Ethics Service Committee of South Central-Oxford C approved the study (reference: 15/SC/0223). This research was undertaken as part of the National Institute of Health Research Health Protection Research Unit (NIHR HPRU) in Blood-Borne and Sexually Transmitted Infections at University College London.

Figure 1. Study design for developing and implementing the bio-behavioral surveillance tool (BBEST) for sexually transmitted infections. MSM: men who have sex with men; STI: sexually transmitted infection.



Setting

Due to dearth of data on contextual drivers of STI among MSM and BC populations [12,13], informed by 2014 GUMCAD data, we decided to implement the BBEST in SHCs that had a high proportion of BC attendees (n=13) and MSM attendees (n=3), referred to henceforth as “BC only study sites” and “MSM only study sites,” respectively. Additional 3 SHCs with high proportions of both BC and MSM attendees were selected and are referred to henceforth as “combined study sites.”

Phase 1: Assessing the Acceptability of the Bio-Behavioral Enhanced Surveillance Tool

Proposed Model of Bio-Behavioral Enhanced Surveillance Tool

We anticipated that the implementation of the BBEST in SHCs would involve the staff offering study envelopes and digital tablets to clinic attendees to participate in a Web-based survey. Participants could also use their personal digital device for survey completion. The survey would be administered using the Snap software (Snap Surveys Ltd, UK) and hosted on a secure remote server.

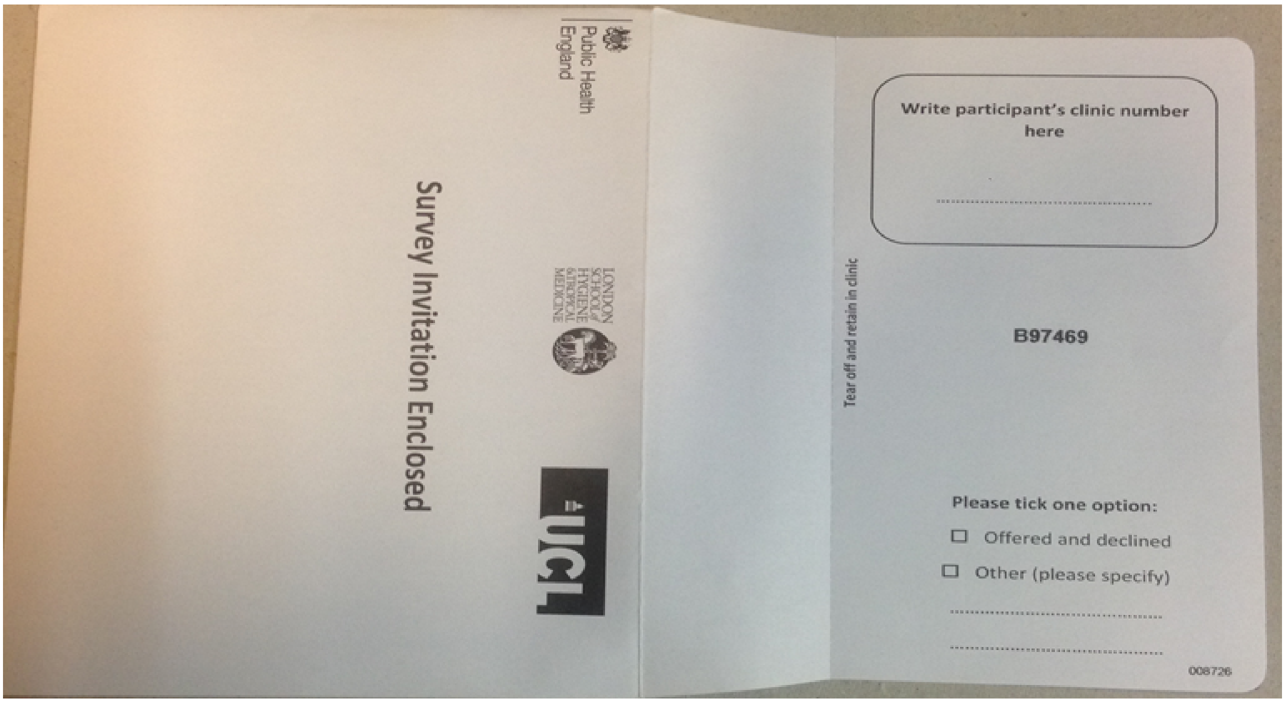
Therefore, internet connectivity would be required in SHCs to enable attendees to log in to the survey and for the data to upload automatically to a remote server. The study team would provide SHCs with study envelopes and tablets that could be remotely deactivated in the event of theft. SHC staff would offer all attendees a study envelope, which would contain a PIS, a card with a survey Web link, and a unique study passcode (USP), which they would use to access the survey. The PIS contained study details including information that only the research team would have access to the survey data, their right to withdraw at any point during or before submitting the survey online, the linkage of their survey responses to the data SHCs routinely collect on STI tests, and results collated by PHE for surveillance purposes. Attendees who could not use digital devices or read English would be ineligible for participation. Each envelope would have a detachable receipt with the same USP (Figure 2). For those who would agree to participate, clinic staff would retain the detachable receipt and write the participant’s clinic number on it to enable linkage of their survey data to their EHR. For attendees who declined to participate, staff would document that on the detachable receipt and retain the entire envelope. Staff would then enter these data into recruitment sheets to be shared monthly with the researchers.

On logging into the survey, participants would again be shown online the same PIS given to them in the study envelope. They would then be asked to give online consent separately for survey participation and linkage of their survey data to the EHR. If they declined to participate, they would exit the survey (Figure 3). If they declined to linkage, they could participate in the survey, but their responses would not be linked to EHR.

Subsequently, they would be screened for eligibility and exit the survey if they reported that they were aged <15 years or ≥15 years but had not had sex in the last year. During online screening, in combined clinics all participants aged ≥16 years who reported having male or male as well as female sexual partner(s) in the last 12 months were directed to the MSM survey and the others were directed to the heterosexual survey. In MSM only clinics, participants who were <16 years, reported having sex only with female partner(s) in last 12 months were excluded.

Subsequently, eligible participants would complete the Web-based survey, which was designed to take approximately 10 to 15 min depending on the sections of the survey that were applicable to them. On completion, researchers could download the participants' survey data from the server and link these to the EHR (for consenting participants only), using the participant's clinic number and USP recorded by the staff, to create a study dataset. We anticipate that the BBEST would be implemented periodically in SHCs or settings that routinely collect GUMCAD data to provide in-depth intelligence on issues of particular public health concern, including STI outbreaks.

Figure 2. Survey invitation envelope: back side with a tear-off tab.



Write participant's clinic number here

....

B97469

Please tick one option:

☐ Offered and declined

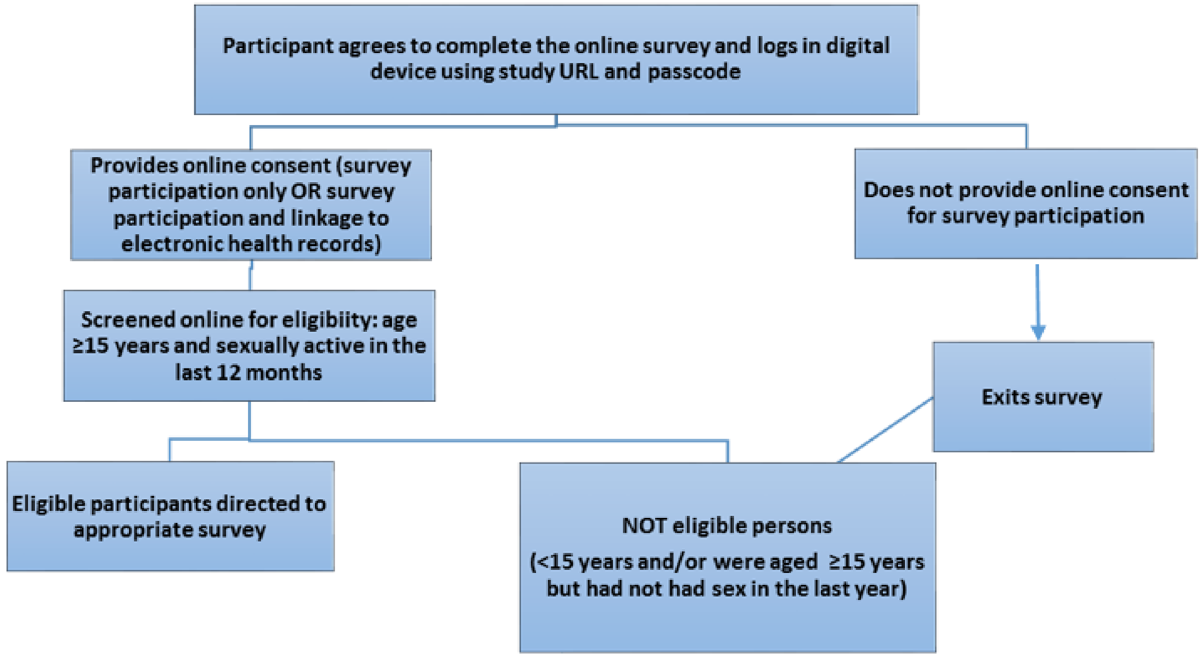
☐ Other (please specify)

....

....

008726

Figure 3. Screening process to identify participants eligible to complete the Web-based survey. EHR: electronic health records.



Qualitative Study

From June 2014 to August 2015, face-to-face or telephone interviews were conducted with 20 staff members recruited from all study sites to assess the acceptability of the proposed BBEST model. A total of 61 MSM were recruited to 1 of the 8 focus group discussions (FGDs) with the help of lesbian, gay, bisexual and trans/sexual health community-based organizations (CBOs) via newsletters and Facebook pages and an MSM geospatial sociosexual networking application [12]. Moreover, 65 BC participants (n=32 men), aged 15-70 years, recruited from SHCs, colleges, and CBOs, participated in 5 FGDs and 31 interviews. All participants were given a PIS containing study details, including information about who would have access to the data and the participants' right to withdraw at any point during or at the end of the interviews and FGDs, and then written informed consent was obtained. Piloted topic guides were used during the FGDs and interviews. Participants were shown printouts of the study envelope (Figure 2), card with a survey Web link and USP, and diagrammatic presentation of the proposed linkage procedure to assess feasibility and acceptability of these processes. All the interviews and FGDs were audio-recorded and transcribed verbatim.

Phase 2: Assessing the Feasibility of Implementing the Bio-Behavioral Enhanced Surveillance Tool

Web-Based Surveys

Two separate Web-based questionnaires were developed, because of differences in STI epidemiology in MSM and BC populations [14], and cognitively tested [15]. One of the surveys was designed to be administered in BC only study sites, and the other in MSM only and combined study sites. Between February and April 2016, the proposed model of BBEST was piloted in 4 SHCs. Subsequently, between May and September 2016, all study sites invited clinic attendees to complete the survey, either in the clinic or at home.

Interviews With Clinic Staff

After completing survey recruitment, short, audio-recorded, semistructured face-to-face/telephone interviews were conducted, between December 2016 and February 2017, with 25 SHC staff from study sites to understand their experiences of implementing the BBEST.

Linkage of Survey Data to Electronic Health Records

Deterministic and probabilistic methods were used to match records in the survey data and GUMCAD using the following key variables: participants' clinic number, age, gender, and clinic attendance date. Probabilistic methods allowed matching of records with erroneous or missing data based on minor discrepancies in age, attendance date, and clinic number. After matching, participants' clinic numbers were dropped to create an anonymous dataset.

Analysis

Phase 1 and phase 2 qualitative data were thematically analyzed using NVivo 11 for Windows (QSR International Pty Ltd, Australia) to assess acceptability of the proposed BBEST model and to examine barriers and facilitators to implementing the BBEST in SHCs, respectively. We used the Framework method to thematically analyze qualitative data [16]. Accordingly, first, we coded the data according to the key areas explored in the topic guides. Subsequently, we coded data for each key topic of interest, for example, all data coded as "digital device for survey," was retrieved, and analyzed to identify themes that summarized participants' common and divergent views concerning the acceptability and feasibility of using digital devices for survey completion, and an index of themes was developed and applied to the qualitative dataset.

The feasibility of implementing the BBEST in phase 2 was assessed quantitatively by examining the recruitment sheets for the number of attendees who were invited to participate in the Web-based survey. The study dataset was examined for the total number of attendees who actually logged in and gave online consent for survey participation and for linkage and the number of eligible participants who completed the survey. The Snap survey software metadata were examined to determine whether a study tablet or other device was used for survey completion. The feasibility of linkage was determined from the number of surveys that were successfully linked to EHR for those who had consented to linkage. Univariable logistic regression was used to examine the association between consent to linkage and the sociodemographic characteristics of survey participants. Representativeness of the survey sample was ascertained by comparing the sociodemographic and sexual health characteristics (only for participants who had consented to and could be linked to their EHR) of the survey participants recruited from the BC only and combined study sites with all the SHC attendees during the study period (data extracted from GUMCAD) using z-test for proportions. The sample recruited from the MSM-only study sites was excluded from this analysis because it was not expected to be representative of "all" SHC attendees due to the study eligibility criteria in these sites (ie, the exclusion of all women and men reporting only female sex partner(s) in the last 12 months; Figure 3). All men identifying as gay/bisexual recruited from all the study sites were compared with all gay/bisexual men accessing these SHCs during the study period to ascertain the representativeness of MSM sample. Stata v13 was used for quantitative data management and analysis.

Results

Acceptability of the Proposed Bio-Behavioral Enhanced Surveillance Tool Model

Perceived Barriers and Facilitators to Implementing Bio-Behavioral Enhanced Surveillance Tool Among Clinic Staff

Internet/Wi-Fi Connectivity

The lack of internet/Wi-Fi connectivity required to administer the Web-based survey and upload the data automatically to a remote server was one of the most commonly perceived barriers by the clinic staff to implementing the BBEST. Staff from clinics located in areas of high deprivation felt that participants may be reluctant to use their own smartphones for survey completion if free Wi-Fi was unavailable in the clinics.

Logistics of Using Tablets for Survey

Clinic staff expressed inability to monitor the security of the tablets because of work pressures and challenges in having dedicated staff members to offer tablets to participants. Although some clinics had separate research staff to help with recruitment, other clinics felt that they would have to depend on regular clinic staff, which was perceived as challenging due to staff cuts that were taking place during our study in several clinics and their high workload. Several research studies being undertaken in the clinic simultaneously were also perceived as a potential barrier. A need to seek help from local clinical research networks (CRNs) who provided temporary staff support for research was identified.

Lack of Experience of Using Digital Devices

Some clinic staff expressed anxiety about using tablets for administering the survey because of their lack of/limited experience of using either tablets or the internet or both. They expressed a need for training to use the tablets and administer the survey.

Documenting Participant's Clinic Number With a Unique Passcode for Linkage to Electronic Health Records

One of the clinics had concerns about sharing the clinic numbers of survey participants who had not agreed for linkage of their survey responses to EHR, but were willing to provide clinic numbers of participants who had given consent to linkage.

Acceptability of Bio-Behavioral Enhanced Surveillance Tool Among Men Who Have Sex With Men and Black Caribbean Participants

Overall, there were few differences in the perceived acceptability of the BBEST among MSM and BC participants, with similar views being expressed by interview and FGD participants.

Using Digital Devices for Survey Completion: Confidentiality

Majority of the participants expressed an ability and willingness to use a digital device to self-complete a Web-based survey because it was considered to be potentially confidential due to immediate online submission of responses post survey completion. Web-based surveys were also perceived to be less embarrassing than a face-to-face paper questionnaire because of the lack of potential for clinic staff to read participants' survey responses. Using a personal digital device compared with a device offered by the clinic for survey completion was preferred because of concerns about applications that may be installed on clinic devices and its impact on confidentiality:

> *Int: And would you be willing to complete this survey on your own device, if you had one, which had access to the internet?*
>
> *IDI_001: I'd feel more comfortable doing it on my own device than something that was given to me.*
>
> *Int: And why do you say that?*
>
> *IDI_001: Because I don't know what else that device has on it, whereas I know what my device has on it. So, like, there are apps that can log key strokes, for example, and stuff like that, so again I'd be trusting that device and that person that gave me that device.*
>
> *Int: Sure.*
>
> *IDI_001: Whereas if it was just a URL and a pass code and I could use my own device, I'd feel much more comfortable doing that.* [BC female interview participant, aged 35 years, Birmingham]

Unlike the BC participants, majority of the MSM were familiar with completing Web-based surveys on their phones and considered it to be a secure and efficient method and had greater preference for single or multiple-choice tick-box questions. However, they were unwilling to download a survey app on their phone. Nevertheless, some participants were not willing to use their personal device for survey completion if they did not have access to free Wi-Fi in the clinic:

> *Respondent 6 Group 5: I've found, because I've done quite a few of these. When you do one, for some reason slightly straight away, once you've done one, this is easy they give you more to do. It almost becomes like great fun.*
>
> *Int: Okay, that's what it feels like?*
>
> *Respondent 6 Group 5: I, I personally find, I'm just talking about me now as an individual, is the multiple choice questions (you) touch on an iPhone, iPad, I love them. I love them because I don't mind.*
>
> *Int: Okay, sure.*
>
> *Respondent 6 Group 5: But I really hate when you have to type in open-ended (answers).* [MSM FGD participant, aged 37 years, Manchester]
>
> *Respondent 5 Group 5: I think you run the risk of people not doing it. You know if it's there in paper form or iPad form or whatever, I'd do it. But I know full well I would leave, probably do the shopping, get the iPad out, life would kick in.* [MSM FGD participant, aged 22 years, Manchester]

Using Digital Devices for Survey Completion: Prior Experience

Participants with experience of completing Web-based surveys perceived it to be a time-saving method because of routing to subsequent questions being informed by their responses to previous questions. However, a few participants expressed an inability to complete a Web-based survey because they did not know how to use the internet, although some were willing to do so if they were shown how to use the digital devices for survey completion:

> *Respondent 1 Group 3: And obviously you can ask questions that, you know…or miss questions. You don't need to ask seventeen questions.* [MSM FGD group participant, aged 36 years, Leeds]
>
> *Respondent 2 Group 3: You only get the ones pertinent to you.* [MSM FGD group participant, aged 46 years, Leeds]

Venue for Survey Completion: Ease and Privacy

Many participants expressed a preference to complete the survey in the clinic because of concerns of getting busy with other things once home. The need to log on to a personal computer once home was perceived to be time-consuming, and not living alone was perceived as a barrier to privacy. However, some participants expressed a preference to complete the survey at home to allow for more considered responses and privacy. Private clinic rooms/booths were preferred for survey completion compared with crowded waiting rooms. Some participants suggested tailoring survey recruitment to the patients flow through the clinic or the layout of services to reduce anxiety about losing their place in the clinic appointment queue:

> *Int: And do you think, if you were to take part in the survey, you'd rather it all happened in the clinic, rather than doing it in your own time at home?*
>
> *IDI-17: Well, not necessarily, it's just when you get home, you've got to make something to eat, you've got to do all this kind of stuff, and then you need to get your computer out and log in and all this kind of stuff, so it's that kind of impetus to do that really.*
>
> *Int: Okay.*
>
> *IDI-17: Whereas if you get given a tablet which is secure obviously, and it's there set up for you ready, then it's a lot kind of easier.* [BC male interview participant, aged 44 years, London]

Linking Survey Responses to Electronic Health Record: Anonymity

Overall, most participants were supportive of the proposed linkage of survey data to EHR because they considered it to "be for something constructive" like improving health care. But some participants perceived it as "too much information gathering on people." Thereby, providing anonymous online consent, separately for linkage to EHR, was perceived to be acceptable. The proposed usage of a USP as opposed to identifiable details to access the survey was considered important to ensure anonymity.

Feasibility of Implementing the Bio-Behavioral Enhanced Surveillance Tool

The above-mentioned findings informed the development of site-specific BBEST models. Temporary staff from the local CRN, who are funded by UK Health Department to provide infrastructural support for patient benefit–related research [17], were arranged to help with the implementation of the study procedures in clinics without dedicated research staff (n=10). In other SHCs, existing and temporary CRN staff (n=2) or preexisting research or administrative staff or both (n=4) were arranged to do so. All staff involved with administering the study were trained by the researchers to implement the standard operating protocol and to use the tablets. Each participating SHC was provided with at least one 3G-enabled iPad and a cable and lock to ensure its security.

Recruitment Offer and Survey Completion Rates

Of the 7367 attendees recorded by SHCs as invited, 6283 (85.28%) agreed to participate (Figure 4); 73.62% (4626/6283) logged in, of whom 21.59% (999/4626) did so using their personal device, and 72.97% (4585/6283) consented to participate in the survey. Moreover, 70.63% (4437/6283) of those who agreed to participate were eligible. However, recruitment success between SHCs varied considerably. Of clinic-specific recruitment targets, 7 SHCs recruited <50% (categorized as "low recruiters"), 5 SHCs recruited between 50% and 80% ("medium recruiters"'), and 4 SHCs recruited >80% ("high recruiters"). Considerably, more clinic attendees were invited to participate in the survey by high recruiters (median 142.23% of the clinic-specific target) compared with low and medium recruiters (57.91% and 74.13%, respectively; Figure 5). Although the proportion of those invited who agreed to participate did not vary between these 3 types of SHCs and ranged between 91% and 93%, the proportion of participants who actually logged in and consented to participate in the survey was higher in medium (85.30%) and high (81.79%) recruiting SHCs than low (40.70%) recruiting SHCs. Similarly, a higher proportion of participants who completed the survey in the clinic in medium (86.19%) and high (83.60%) recruiting SHCs than low (33.12%) recruiting SHCs used a study tablet to log into the survey. There was no difference in the proportion of participants who were eligible for the survey across these groups (range: 92%-97%).

Barriers and Facilitators to Implementing the Bio-Behavioral Enhanced Surveillance Tool

Phase 2 staff interviews highlighted that SHC with a culture of valuing and prioritizing research and innovation, championed by senior clinicians, was an important factor for successfully implementing the BBEST. Having enthusiastic staff members, especially with dedicated roles and time to implement the study, maximized the recruitment. However, structural changes in clinical services, including a London-wide reorganization of sexual health, negatively affected staff morale and their engagement with the study, and in some SHCs, it led to frequent staff turnover. This affected recruitment with subsequent staff receiving limited training in study procedures, resulting, for example, in one clinic not documenting participants' clinic numbers. Nevertheless, despite initial concerns, the majority of staff involved with implementing the survey became familiar and gained confidence with the use of tablets. However, some clinics offered tablets to participants only if they had dedicated staff to monitor them because of concerns of theft or damage. One tablet was stolen from a locked cabinet in a staff-only access area, and in 2 clinics, the lack of ability to secure tablets to immovable objects prevented their use. The limited number of tablets available per clinic for administering the survey restricted their ability to recruit multiple participants simultaneously. However, in clinics that had Wi-Fi, the option for participants to complete the survey using their own digital device was considered a facilitator for recruitment. Some staff felt overburdened with the requirement to record participants' clinic number and sending the recruitment sheet monthly to the research team. Nonetheless, they appreciated the real-time feedback from the study team on survey completion rates, which enabled them to promote completion in the clinic, as opposed to home, to address lower home completion rates.

Feasibility of Linking Survey Data to Electronic Health Records

Altogether, 91.19% (4046/4437) of the eligible participants consented to link their survey data to EHR (Figure 4). This did not differ by age or gender but was higher among gay/bisexual than heterosexual men (95.50%, 722/756 vs 88.31%, 1073/1215; *P*<.003) and lower among BC than white participants (87.25%, 568/651 vs 93.89%, 2181/2323; *P*<.002). Of those participants who had consented to linkage, 88.88% (3596/4046) of surveys were successfully linked to EHR (80.98%, 3593/4437, of eligible participants). Moreover, 83.49% (3000/3593) of these records matched on all the variables, with the remaining matches differing slightly in at least 1 variable, including 3.23% with typographical errors in the USP. In addition, 34.7% (156/450) surveys could not be linked to EHR because the staff did not record either the USP or the participants' clinic number. The remaining surveys could not be linked either because of errors in the clinic number that could not be resolved using the probabilistic linkage algorithm or a mismatch between the participant's clinic number used in the clinic compared with that submitted to GUMCAD.

Figure 4. Feasibility of implementing the Bio-Behavioral Enhanced Surveillance Tool by population of interest. MSM: men who have sex with men.

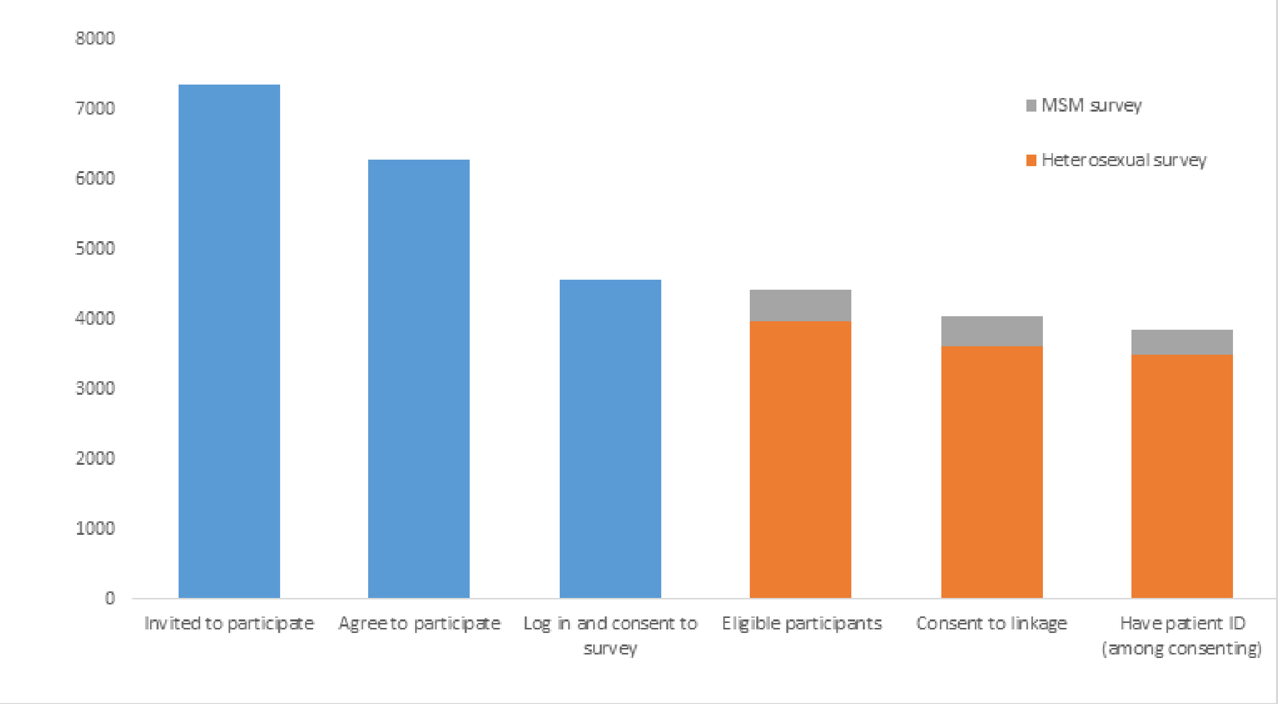


Figure 5. Recruitment cascade among clinics grouped by recruitment success.

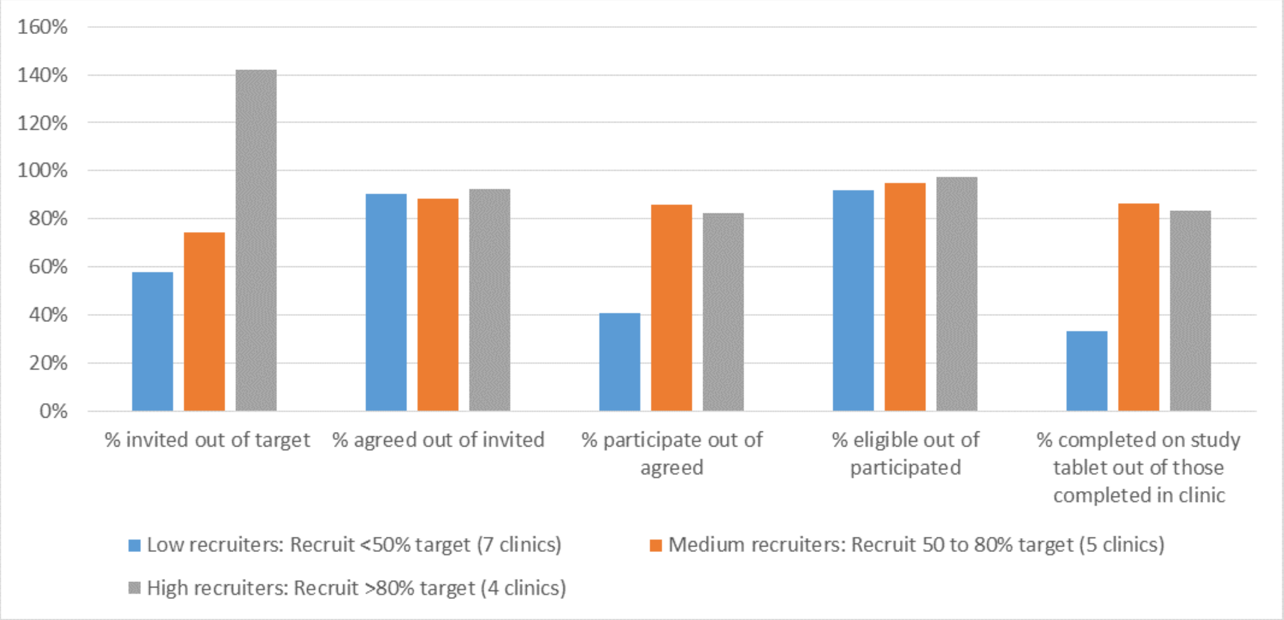


Table 1. Comparison of sociodemographic characteristics of survey participants with all clinic attendees during the study period.

Sociodemographic characteristics and indicators of sexual health	GUMCAD data on attendees in clinics administering only heterosexual survey (N=97,054)		Survey population in clinics administering only heterosexual survey (N=4184)		*P* value (Z test for proportions comparing non-missing)
	n (%)	Percentage excluding unknown/missing variables	n (%)	Percentage excluding unknown/missing variables	
Gender					
Male	41,609 (42.87)	42.87	1722 (41.16)	41.16	.03
Female	55,411 (57.09)	57.09	2447 (58.48)	58.48	.08
Trans/other[a]	Not available	Not available	15 (0.36)	0.36	Not calculated
Unknown/missing	34 (0.00)	—[b]	0 (0.00)	0.00	—
Sexual orientation (of all men)	**N=41,609**		**N=1722**		
Gay/bisexual	9597 (23.06)	23.90	519 (30.14)	31.52	<.001
Heterosexual	30,585 (73.51)	76.10	1165 (67.65)	70.61	<.001
Unknown	1427 (3.43)	—	73 (4.24)	—	—
Age, in years					
<25	30,034 (30.94)	30.94	1,599 (38.21)	38.21	<.001
≥25	67,020 (69.05)	69.05	2585 (61.78)	61.78	<.001
Unknown/missing	0 (0.00)	—	0 (0.00)	—	—
Ethnicity					
White	49,409 (50.91)	56.00	2092 (50.00)	51.20	<.001
Black African	7960 (8.19)	9.01	418 (9.99)	10.2	<.001
Black Caribbean	8368 (8.62)	9.51	640 (15.29)	15.7	<.001
Black other	4441 (4.60)	5.01	42 (1.00)	1.00	<.001
Mixed	6549 (6.78)	7.41	370 (8.80)	9.11	<.001
Asian	8224 (9.29)	9.29	400 (9.62)	9.81	.32
Other ethnicities	3288 (3.38)	3.38	125 (2.98)	2.98	.25
Unknown/missing	8815 (9.08)	—	97 (2.31)	—	—
Sexual health outcomes using total linked data for clinics that offered heterosexual survey[c]			**N=3447**		
Sexual health screen on day of clinic attendance[d]					
Yes	74,217 (76.52)	76.52	2552 (73.98)	73.98	<.001
No	22,837 (23.53)	23.53	895 (25.96)	25.96	<.001
Diagnosed with an acute STI[e] on the day of clinic attendance[c,f]					
Yes	14,240 (14.67)	14.67	585 (16.97)	16.97	<.001
No	82,814 (85.32)	85.32	2862 (83.03)	83.03	<.001

[a] Trans not currently recorded on GUMCAD surveillance.

[b] Indicates "not applicable".

[c] There is no “missing” data for these categories as there is no requirement to code when there is no STI screen or STI diagnosis.

[d] Sexual health screen—one of the following test combinations: chlamydia and gonorrhea; chlamydia, gonorrhea, and syphilis; syphilis and HIV; and chlamydia, gonorrhea, syphilis, and HIV.

[e] STI: sexually transmitted infections.

[f] Acute STI—any of chlamydia, gonorrhea, anogenital herpes (first episode), anogenital herpes (first episode), HIV, infectious syphilis, pelvic inflammatory disease/epididymitis, non-specific genital infections, chancroid, lymphogranuloma venereum, donovanosis, trichomoniasis, scabies, pediculosis pubis, molluscum contagiosum, mycoplasma genitalium, shigella, hepatitis A, hepatitis B, and hepatitis C.

Table 2. Comparison of sociodemographic characteristics of men who have sex with men (MSM) who participated in the survey compared with all MSM attending all the study sites during the study period.

Sociodemographic characteristics and indicators of sexual health	GUMCAD data (N=11,180)		Survey population (N=751)		*P* value (Z test for proportions comparing non-missing)
	n (%)	Percentage excluding unknown/missing variables	n (%)	Percentage excluding unknown/missing variables	
Age, in years					
<25	1776 (15.88)	15.88	167 (22.2)	22.2	<.001
≥25	9404 (84.11)	84.11	584 (77.8)	77.8	<.001
Unknown/missing	0 (0.00)	0.00	0 (0.0)	0.0	—[a]
Ethnicity					
White	7671 (68.61)	75.00	551 (73.3)	74.3	.70
Black African	226 (2.02)	2.02	15 (2.0)	2.0	.74
Black Caribbean	350 (3.13)	3.40	37 (4.9)	5.0	.03
Black other	132 (1.80)	1.31	2 (0.3)	0.3	.02
Mixed	487 (4.35)	4.80	50 (6.7)	6.7	.02
Asian	876 (7.83)	8.61	62 (8.3)	8.3	.85
Other ethnicities	486 (4.34)	4.81	24 (3.2)	3.2	.06
Unknown/missing	952 (8.52)	–	10 (1.33)	–	–
Sexual health outcomes using total linked data for clinics that offered heterosexual survey[b]			**N=571**		
Sexual health screen on day of clinic attendance[c]					
Yes	9051 (80.89)	80.89	406 (71.1)	71.1	<.001
No	2129 (19.04)	19.04	165 (28.9)	28.9	<.001
Diagnosed with an acute STI[d] **on the day of clinic attendance**[b, e]					
Yes	2203 (19.70)	19.70	116 (20.3)	20.3	.72
No	8977 (80.29)	80.29	455 (79.7)	79.7	.72

[a] Indicates "not applicable".

[b] There is no “missing” for these categories as there is no requirement to code when there is no STI screen or STI diagnosis.

[c] Sexual health screen—one of the following test combinations: chlamydia and gonorrhea; chlamydia, gonorrhea, and syphilis; syphilis and HIV; and chlamydia, gonorrhea, syphilis, and HIV.

[d] STI: sexually transmitted infections.

[e] Acute STI—any of chlamydia, gonorrhea, anogenital herpes (first episode), anogenital herpes (first episode), HIV, infectious syphilis, pelvic inflammatory disease/epididymitis, nonspecific genital infections, chancroid, lymphogranuloma venereum, donovanosis, trichomoniasis, scabies, pediculosis pubis, molluscum contagiosum, mycoplasma genitalium, shigella, hepatitis A, hepatitis B, and hepatitis C.

Representativeness of the Survey Sample

Compared with all the SHC attendees accessing BC only and combined study sites during the study period (N=97,054), of the 4184 survey participants recruited from these sites a higher proportion were aged <25 years (38.21%, 1599/4184 vs 30.94%, 30034/97,054; *P*<.001) and were BC (15.29%, 640/4184 vs 8.62%, 8368/97,054; *P*<.001). There was no overall difference in recruitment by sex, but men recruited from these study sites were more likely to identify as gay/bisexual (42.92%, 519/1209 vs 23.06%, 9597/41,609; *P*<.001; Table 1). Responses of 82.39% (3447/4184) of participants’ who were attending clinics that offered the heterosexual survey were linked to their EHR. Compared with all the SHC attendees, a slightly lower proportion of this survey sample had had a same-day sexual health screen during that clinic visit (74.97%, 2552/3447 vs 76.52%, 74,217/97,054; *P*<.001), but a slightly higher proportion of them were diagnosed with acute STIs during that clinic visit (16.97%, 585/3447 vs 14.67%, 14,240/97,054; *P*<.001).

As shown in Table 2, compared with all gay/bisexual identifying men attending SHCs during the study period (N=11,180), a higher proportion of men in the survey sample were <25 years (22.2%, 167/751 vs 15.88%, 1776/11,180; *P*<.001) and were BC (4.9%, 37/751 vs 3.13%, 350/11,180; *P*=.025), but a lower proportion had had a same-day sexual health screen (71.1%, 406/571 vs 80.89%, 9051/11,189; *P*<.001). There was no difference in the proportion who were diagnosed with an acute

STI during that clinic visit (20.3%, 116/571 vs 19.70%, 2203/11,180 [*P*=.721]).

Discussion

Key Findings

Our findings show that the BBEST is largely acceptable to SHC attendees and staff, and it is feasible to implement in SHCs across England. Specifically, the SHC attendees at greatest STI risk participated in the self-administered Web-based surveys using digital devices and consented to linkage. Linking survey data to the EHR was also feasible. However, a lack of resources dedicated to delivering the BBEST was a barrier to its implementation in some SHCs.

Strengths and Limitations

Study sites were purposively selected and thus are not representative of SHCs in England. Moreover, the survey offer rates varied considerably across sites increasing the likelihood of recruitment bias. We are unable to fully assess the representativeness of our sample because of the lack of data on decliners, attendees who agreed to participate but did not log in, and those who logged in but did not consent to participate. The representativeness analysis shows that, overall, groups at greater STI risk are overrepresented in the survey, for example, participants aged <25 years, gay/bisexual men, and BC participants. Nevertheless, the proportion of gay/bisexual identifying men who had an acute STI diagnosis on the day of clinic attendance was similar among those in the survey and in the clinic population, highlighting the similar STI risk profile of these men.

Comparison With Other Studies

Overall, the response rate among those invited to participate was 62.23% (4585/7367); however, survey offer rates between SHCs varied enormously. Similar to our study, this interclinic variation was observed in previous clinic surveys conducted using pen and paper, with response rates among attendees across clinics varying from 41.0% to 70.1% in one study [18] and from 24.9% to 76.1% in another [19]. In both these studies, and as we observed, this variation was attributed to differences in staff's commitment to, and enthusiasm for, the study and resources available within SHCs for research [18,20]. In our study, CRNs were unable to provide support for recruitment to some SHCs because of a high demand on their resources from multiple studies; however, the majority of sites with CRN support met their recruitment target. Although increasing the CRN support available to clinics may improve their ability to participate in research and thus their response rate, sometimes this may not compensate for "research fatigue," among clinic staff and clinic attendees from participating in multiple studies taking place in the clinic, which negatively impacts the response.

Compared with another study, a slightly higher proportion of participants in our study consented to linkage of their survey data to their EHR (91.2% vs 84.0%) [20]. This could potentially be due to the cocreation of the BBEST with the involvement of the CAG and SHC attendees and staff. The feasibility of linking the survey data to EHR in our study was high. Similar to findings of a previous study, linkage was unsuccessful in a handful of cases due to minor errors [18]; nevertheless, using a probabilistic method increased match sensitivity.

Implications for Practice

Our findings suggest that the BBEST can be implemented to collect detailed behavioral data on factors influencing STI risk among SHC attendees, especially among populations at greatest STI risk in the event of outbreaks or periodically in response to significant public health concerns. Furthermore, the BBEST could be implemented in other settings such as hospitals, general practice, and with other populations at STI/HIV risk to strengthen interpretation of existing surveillance data. However, the impact of increasing private health care providers [21,22] on the availability of resources for implementing the BBEST is as yet unknown. Moreover, CRNs provide support only for research studies and not for surveillance activities. Therefore, extension of CRN support to surveillance activities could enhance implementation of BBEST. Although Web-based self-administered surveys were acceptable to SHC attendees, several SHCs did not have a reliable internet connection/Wi-Fi, highlighting the need to rapidly scale up the implementation of plans for digitizing the NHS for patient benefit [23]. Survey software that facilitates real-time data uploading from digital devices to a secure remote server and digital devices that can be remotely deactivated should be used to enhance data security and participants' confidentiality. PHE does not need SHC attendees' consent to collect clinical and behavioral data for public health monitoring and response work under section 251 of the UK NHS Act of 2006, although there are strict regulations for doing so [24]. However, obtaining informed consent is considered a norm in health surveys [25] and is particularly important in the context of linking surveillance data to survey data. Our study shows that informed consent can be obtained online, anonymously, and explicitly for survey participation and for linking it to the EHR. In conclusion, in an era of reduced sexual health budgets [22,26] and the limitations of existing routine STI surveillance methods, implementing a BBEST could enable timely collection of detailed behavioral data to better inform effective commissioning of health promotion and STI prevention strategies. However, feasibility of implementing BBEST could be influenced by the availability of resources.

Acknowledgments

The authors thank all the study participants and staff of SHCs involved in this study. The authors also thank all the community-based organizations who helped them to recruit participants for the qualitative study. The authors acknowledge the members of the National Institute of Health Research Health Protection Research Unit (NIHR HPRU) in Blood Borne and Sexually Transmitted Infections Steering Committee: Caroline Sabin, Anthony Nardone, Catherine Mercer, Gwenda Hughes, Greta Rait, Jackie Cassell, William Rosenberg, Tim Rhodes, Kholoud Porter, and Samreen Ijaz and the members of Theme A of the NIHR HPRU in Blood

Borne and Sexually Transmitted Infections Steering Committee: Catherine Mercer, Gwenda Hughes, Hamish Mohammed, Jackie Cassell, Fiona Burns, Makeda Gerressu, Jonathan Elford, David Phillips, Gary Brook, Nicola Low, Anthony Nardone, Sarika Desai, Adamma Aghaizu, Alison Rodgers, and Paul Crook. The authors would like to thank Catherine Griffiths and Catherine Aicken for their help with the systematic review; Victoria Gilbart, Jessica Datta, and Emma Garnett for their help with qualitative data collection; and Lorna Sutcliffe and Jessica Datta for their help with coding qualitative data. The research was funded by the NIHR HPRU in Blood Borne and Sexually Transmitted Infections at UCL in partnership with Public Health England (PHE) and in collaboration with the London School of Hygiene and Tropical Medicine. SW and DR are funded by the NIHR HPRU in Blood Borne and Sexually Transmitted Infections at UCL in partnership with PHE and in collaboration with the London School of Hygiene and Tropical Medicine.

Authors' Contributions

SW, DR, and PB set up the study, managed by PW, GH, and CM. SW and DR secured ethics and R&D permissions, and SW coordinated and managed the implementation of the study in all study sites, with support from PW, GH, and CM. SW undertook the systematic review; SW and DR undertook qualitative data collection and data analyses; SW, DR, and PB oversaw the delivery of the patient survey in GUM clinics; PB undertook the linkage of the patient survey data to GUMCAD. SW and PB analyzed the quantitative data. CM and GW secured funding from the National Institute for Health Research for the Health Research Health Protection Research Unit (NIHR HPRU) in Blood Borne and Sexually Transmitted Infections at University College London in partnership with Public Health England (PHE), in collaboration with London School of Hygiene & Tropical Medicine. All authors contributed to the drafting of the paper and approved the final version. The views expressed are those of the authors and not necessarily those of the NHS, the NIHR, the Department of Health, or Public Health England.

Conflicts of Interest

None declared.

References

1. Mohammed H, Hughes G, Fenton KA. Surveillance systems for sexually transmitted infections: a global review. Curr Opin Infect Dis 2016 Feb;29(1):64-69. [doi: 10.1097/QCO.0000000000000235] [Medline: 26658655]
2. Newman L, Rowley J, Vander Hoorn S, Wijesooriya NS, Unemo M, Low N, et al. Global estimates of the prevalence and incidence of four curable sexually transmitted infections in 2012 based on systematic review and global reporting. PLoS One 2015;10(12):e0143304 [FREE Full text] [doi: 10.1371/journal.pone.0143304] [Medline: 26646541]
3. Furegato M, Chen Y, Mohammed H, Mercer CH, Savage EJ, Hughes G. Examining the role of socioeconomic deprivation in ethnic differences in sexually transmitted infection diagnosis rates in England: evidence from surveillance data. Epidemiol Infect 2016 Nov;144(15):3253-3262. [doi: 10.1017/S0950268816001679] [Medline: 27511704]
4. Public Health England. 2017. What we do URL: https://www.gov.uk/government/organisations/public-health-england/about [WebCite Cache ID 6tEGY469y]
5. Public Health England. Sexually Transmitted Infections and Chlamydia Screening in England, 2016 URL: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/617025/Health_Protection_Report_STIs_NCSP_2017.pdf [WebCite Cache ID 6tEGnoRgZ]
6. Public Health England. Genitourinary medicine clinic activity dataset (GUMCADv3): pilot, 2015 URL: https://www.gov.uk/guidance/genitourinary-medicine-clinic-activity-dataset-gumcadv3-pilot [WebCite Cache ID 6tEGyrzlH]
7. UNAIDS/WHO Working Group on Global HIV/AIDS & STI Surveillance. Geneva: UNAIDS; 2000. Guidelines for second generation HIV surveillance URL: http://apps.who.int/iris/bitstream/10665/66297/1/WHO_CDS_CSR_EDC_2000.5.pdf [WebCite Cache ID 6tEJSl3nu]
8. Gilbart VL, Simms I, Jenkins C, Furegato M, Gobin M, Oliver I, et al. Sex, drugs and smart phone applications: findings from semistructured interviews with men who have sex with men diagnosed with Shigella flexneri 3a in England and Wales. Sex Transm Infect 2015 Dec;91(8):598-602. [doi: 10.1136/sextrans-2015-052014] [Medline: 25921020]
9. Richens J, Copas A, Sadiq ST, Kingori P, McCarthy O, Jones V, et al. A randomised controlled trial of computer-assisted interviewing in sexual health clinics. Sex Transm Infect 2010 Aug;86(4):310-314. [doi: 10.1136/sti.2010.043422] [Medline: 20551234]
10. Johnson AM, Copas AJ, Erens B, Mandalia S, Fenton K, Korovessis C, et al. Effect of computer-assisted self-interviews on reporting of sexual HIV risk behaviours in a general population sample: a methodological experiment. AIDS 2001 Jan 5;15(1):111-115. [Medline: 11192852]
11. Langhaug LF, Sherr L, Cowan FM. How to improve the validity of sexual behaviour reporting: systematic review of questionnaire delivery modes in developing countries. Trop Med Int Health 2010 Mar;15(3):362-381 [FREE Full text] [doi: 10.1111/j.1365-3156.2009.02464.x] [Medline: 20409291]
12. Datta J, Reid D, Hughes G, Mercer CH, Wayal S, Weatherburn P. Places and people: the perceptions of men who have sex with men concerning STI testing: a qualitative study. Sex Transm Infect 2018 Feb;94(1):46-50 [FREE Full text] [doi: 10.1136/sextrans-2016-052983] [Medline: 28778980]

13. Wayal S, Hughes G, Sonnenberg P, Mohammed H, Copas AJ, Gerressu M, et al. Ethnic variations in sexual behaviours and sexual health markers: findings from the third British National Survey of Sexual Attitudes and Lifestyles (Natsal-3). Lancet Public Health 2017 Oct;2(10):e458-e472 [FREE Full text] [doi: 10.1016/S2468-2667(17)30159-7] [Medline: 29057382]
14. Hughes G, Field N. The epidemiology of sexually transmitted infections in the UK: impact of behavior, services and interventions. Future Microbiol 2015;10(1):35-51. [doi: 10.2217/fmb.14.110] [Medline: 25598336]
15. Willis G. Cognitive Interviewing: A How To Guide?. USA: Research Triangle Institute; 1999.
16. Gale NK, Heath G, Cameron E, Rashid S, Redwood S. Using the framework method for the analysis of qualitative data in multi-disciplinary health research. BMC Med Res Methodol 2013 Sep 18;13:117 [FREE Full text] [doi: 10.1186/1471-2288-13-117] [Medline: 24047204]
17. United Kingdom Clinical Research Collaboration. Clinical Research Networks 2005 URL: http://www.ukcrc.org/research-infrastructure/clinical-research-networks/clinical-research-networks-in-england/ [WebCite Cache ID 6tEHHtHJN]
18. Mercer CH, Sutcliffe L, Johnson AM, White PJ, Brook G, Ross JD, et al. How much do delayed healthcare seeking, delayed care provision, and diversion from primary care contribute to the transmission of STIs? Sex Transm Infect 2007 Aug;83(5):400-405 [FREE Full text] [doi: 10.1136/sti.2006.024554] [Medline: 17475683]
19. Mercer CH, Aicken CR, Tanton C, Estcourt CS, Brook MG, Keane F, et al. Serial monogamy and biologic concurrency: measurement of the gaps between sexual partners to inform targeted strategies. Am J Epidemiol 2013 Jul 15;178(2):249-259. [doi: 10.1093/aje/kws467] [Medline: 23801013]
20. Aicken CR, Cassell JA, Estcourt CS, Keane F, Brook G, Rait G, et al. Rationale and development of a survey tool for describing and auditing the composition of, and flows between, specialist and community clinical services for sexually transmitted infections. BMC Health Serv Res 2011 Feb 09;11:30 [FREE Full text] [doi: 10.1186/1472-6963-11-30] [Medline: 21306604]
21. Clarke J, Carlin E. Sexual health is another commissioning crisis. Br Med J 2015 Jan 14;350:h149. [Medline: 25588951]
22. Robertson RW, Thompson J, Charles A. Kingsfund. London, UK: The King's Fund; 2017. Understanding NHS financial pressures. How are they affecting patient care? URL: https://www.kingsfund.org.uk/sites/default/files/field/field_publication_file/Understanding%20NHS%20financial%20pressures%20-%20full%20report.pdf [accessed 2017-09-04] [WebCite Cache ID 6tEIYavTw]
23. National Information Board, Department of Health. 2014. Personalised health and care 2020: a framework for action URL: https://www.gov.uk/government/publications/personalised-health-and-care-2020/using-data-and-technology-to-transform-outcomes-for-patients-and-citizens [WebCite Cache ID 6tEJ3w6AL]
24. Public Health England. Personal information charter URL: https://www.gov.uk/government/organisations/public-health-england/about/personal-information-charter [WebCite Cache ID 6ybLyqM8k]
25. World Health Organisation. Geneva: World Health Organisation; 2017 Sep 04. Guidelines on ethnical issues in public health surveillance URL: http://apps.who.int/iris/bitstream/10665/255721/1/9789241512657-eng.pdf [WebCite Cache ID 6tEHalnTd]
26. HM Treasury and The Rt Hon George Osborne. 2015. Chancellor announces 4 and a half billion pounds of measures to bring down debt URL: https://www.gov.uk/government/news/chancellor-announces-4-billion-of-measures-to-bring-down-debt [WebCite Cache ID 6tEHR2d8F]

Abbreviations

BBEST: bio-behavioral enhanced surveillance tool
BC: black Caribbean
CAG: Community Advisory Group
CASI: computer-assisted self-interview
CBOs: community-based organizations
CRN: clinical research network
EHR: electronic health records
FGD: focus group discussions
MSM: men who have sex with men
PIS: participant information sheet
SHC: sexual health clinics
STI: sexually transmitted infections
USP: unique study passcode

Edited by P Mee; submitted 21.09.17; peer-reviewed by J Elmes, M Puttaroo; comments to author 08.12.17; accepted 19.02.18; published 04.05.18.

Please cite as:
Wayal S, Reid D, Blomquist PB, Weatherburn P, Mercer CH, Hughes G
The Acceptability and Feasibility of Implementing a Bio-Behavioral Enhanced Surveillance Tool for Sexually Transmitted Infections in England: Mixed-Methods Study
JMIR Public Health Surveill 2018;4(2):e52
URL: http://publichealth.jmir.org/2018/2/e52/
doi:10.2196/publichealth.9010
PMID:29728348

Original Paper

Clinical Relation Extraction Toward Drug Safety Surveillance Using Electronic Health Record Narratives: Classical Learning Versus Deep Learning

Tsendsuren Munkhdalai[1], PhD; Feifan Liu[1], PhD; Hong Yu[2,3], FACMI, PhD

[1]Department of Quantitative Health Sciences, University of Massachusetts Medical School, Worcester, MA, United States
[2]Department of Computer Science, University of Massachusetts Lowell, Lowell, MA, United States
[3]The Bedford Veterans Affairs Medical Center, Bedford, MA, United States

Corresponding Author:
Hong Yu, FACMI, PhD
Department of Computer Science
University of Massachusetts Lowell
1 University Ave
Lowell, MA, 01854
United States
Phone: 1 9789343620
Fax: 1 9789343551
Email: hong_yu@uml.edu

Abstract

Background: Medication and adverse drug event (ADE) information extracted from electronic health record (EHR) notes can be a rich resource for drug safety surveillance. Existing observational studies have mainly relied on structured EHR data to obtain ADE information; however, ADEs are often buried in the EHR narratives and not recorded in structured data.

Objective: To unlock ADE-related information from EHR narratives, there is a need to extract relevant entities and identify relations among them. In this study, we focus on relation identification. This study aimed to evaluate natural language processing and machine learning approaches using the expert-annotated medical entities and relations in the context of drug safety surveillance, and investigate how different learning approaches perform under different configurations.

Methods: We have manually annotated 791 EHR notes with 9 named entities (eg, medication, indication, severity, and ADEs) and 7 different types of relations (eg, medication-dosage, medication-ADE, and severity-ADE). Then, we explored 3 supervised machine learning systems for relation identification: (1) a support vector machines (SVM) system, (2) an end-to-end deep neural network system, and (3) a supervised descriptive rule induction baseline system. For the neural network system, we exploited the state-of-the-art recurrent neural network (RNN) and attention models. We report the performance by macro-averaged precision, recall, and F1-score across the relation types.

Results: Our results show that the SVM model achieved the best average F1-score of 89.1% on test data, outperforming the long short-term memory (LSTM) model with attention (F1-score of 65.72%) as well as the rule induction baseline system (F1-score of 7.47%) by a large margin. The bidirectional LSTM model with attention achieved the best performance among different RNN models. With the inclusion of additional features in the LSTM model, its performance can be boosted to an average F1-score of 77.35%.

Conclusions: It shows that classical learning models (SVM) remains advantageous over deep learning models (RNN variants) for clinical relation identification, especially for long-distance intersentential relations. However, RNNs demonstrate a great potential of significant improvement if more training data become available. Our work is an important step toward mining EHRs to improve the efficacy of drug safety surveillance. Most importantly, the annotated data used in this study will be made publicly available, which will further promote drug safety research in the community.

(JMIR Public Health Surveill 2018;4(2):e29) doi:10.2196/publichealth.9361

KEYWORDS

medical informatics applications; drug-related side effects and adverse reactions; neural networks; natural language processing; electronic health records

Introduction

Background and Significance

Prescription drug safety represents a major public health concern [1]. An adverse drug event (ADE) is “an injury resulting from medical intervention related to a drug” [2]. ADEs are common and occur in approximately 2-5% of hospitalized adult patients [2-5]. Each ADE is estimated to increase the length of a hospital stay by more than 2 days and hospital cost by more than US $3200 [4,6]. Severe ADEs rank among the top 4 or 6 leading causes of death in the United States [7]. Prevention, early detection, and mitigation of ADEs could save both lives and resources [6,8,9].

Due to the limited number of participants and inclusion or exclusion criteria reflecting specific subject characteristics, premarketing randomized clinical trials frequently miss ADEs [1], and thus, postmarketing drug safety surveillance [10] is vitally important for health care and patient safety. The Food and Drug Administration (FDA) maintains an adverse event reporting system called the Food and Drug Administration Adverse Event Reporting System for postmarketing safety surveillance, but it faces challenges including underreporting [11,12] and missing important patterns of drug exposure [13]. Other resources have been shown to be useful for identifying ADEs, including biomedical literature [14] and social media [15-18]. However, biomedical literature has been shown to identify mostly a limited set of rare ADEs [19]. Social media has its own challenges, such as missing important drug exposure patterns and generalizing system to deal with data heterogeneity [17].

It is well known that electronic health records (EHRs) contain rich ADE information and are an important resource for drug safety surveillance [2,20,21]. Since 2009, the FDA has invested in facilitating the use of routinely collected EHR data to perform active surveillance of the safety of marketed medical products [22]. Existing ADE-targeted observational studies have focused on structured EHR data for obtaining ADE information [23-25]; however, ADEs are often buried in the EHR narratives and not recorded in structured data. Manual abstraction of data from EHR notes [5,26] remains a costly and significant impediment to drug safety surveillance research. Exploring natural language processing (NLP) approaches for efficient, accurate, and automated ADE detection can provide significant cost and logistical advantages over manual chart review or voluntary reporting.

Mining Clinical Narratives for ADE Detection

Quite a few NLP approaches have been explored for mining ADE information from unstructured data of the aforementioned sources, such as biomedical literature [27,28], social media [29], FDA event reporting system narratives [30], and EHRs [31-40]. The 2009 i2b2 (Informatics for Integrating Biology and the Bedside) medication challenge [41] and the 2010 i2b2 relation challenge [42] plays an important role to promote methodology advancement in this field. Existing studies are limited to detect only on the document level by identifying discharge summaries that contains ADE [31], or mainly focus on detecting entities representing relevant events (eg, adverse events and medication events) [32,33,43], or deal with only intrasentential relations [42], or identify relations purely based on statistical association analysis among drug and outcome concepts, which are recognized by mapping free clinical text onto medical terminology [37-40]. Henriksson et al [35] explored traditional random-forest algorithm to identify relations between drugs and disorders (or findings) on Swedish clinical notes, and reported that the intersentential relations are challenging and hard to detect.

Recently, deep learning with neural networks has received increasing attention in NLP tasks [44,45], and for relation extraction, the state-of-the-art systems are based on 2 networks: recurrent neural networks (RNNs) [46,47] and convolutional neural networks (CNNs) [48], and an end-to-end relation extraction model [49] obtained competitive performance on several datasets. So far, there is less related work on evaluating deep learning methods on ADE relation extraction. Li et al [50] proposed a bidirectional LSTM to extract ADE relations from biomedical literature. As the model is dependent on the parsing of a sentence, it is difficult to apply that on clinical notes which contain more abbreviations and ungrammatical language expressions. In clinical domain, Lv et al [51] combined autoencoder with conditional random fields, and Sahu et al [52] proposed a domain invariant CNNs for ADE extraction on the i2b2 data. All the 3 studies are limited to extract relations within 1 sentence.

Objective

In this study, we investigate ADE-relevant relation extraction on both intra- and intersentential settings. To this end, we have built a benchmark corpus consisting of clinical notes where medical concepts related to ADE and their relations were annotated via a manual chart review. Then, we experimented with 3 supervised machine learning approaches for ADE relation identification from clinical notes. The first approach is based on rule induction, which is similar to supervised descriptive rule induction [53] but is relatively simple. Rules for each relation type are automatically induced based on the corresponding descriptive statistics obtained from the training data, and then those rules are used to classify new entity pairs. Our second approach uses a classical support vector machines (SVM)-based machine learning model. Our third approach is based on deep learning neural networks, which explore RNNs with attention mechanisms. In addition to benchmark the overall performance, we empirically analyzed how well deep learning models are in terms of recognizing long-distance relations, and how the training data size affects learning performance on clinical data. Compared with previous studies, the main contributions of this work are as follows:

- We build a new annotated benchmark corpus of EHR notes for ADE information extraction. Compared with the existing i2b2 data, this corpus contains much richer annotations related to ADE research, for example, all the medications are profiled with attributes enabling ADE connected to a specific dose of medication (note that many ADEs are caused by high dosage); severity concepts are also annotated and associated with ADEs.
- The annotated data in this study will be shared with the community to further promote research for drug safety surveillance.
- It is the first attempt to investigate and evaluate modeling 7 heterogeneous clinical relations in a single framework: relations between medication and its attributes, relations between ADE and its severity, relations between medication and ADE, and relations between medication and indication.
- We explored RNNs and attention mechanisms for clinical relation extraction beyond sentence boundaries, and investigate how the length between two entities affects the performance for different learning models. To our knowledge, this is the first study of applying deep learning approaches on both inter- and intrasentential relation extraction using EHR data.

Methods

Data Annotation

The annotated corpus contains 791 English EHR notes from cancer patients, which were randomly sampled from people who have been diagnosed with hematological malignancy and have drug exposure to one or more of the 12 cancer drugs of interest, including Romidepsin, Rituximab, Brentuximab vedotin, Ponatib, Carfilzomib. All the notes are longitudinal and no note type filtering was performed. We manually annotated 8 named entities and 7 relation types among them: *Dosage-Medication, Route-Medication, Frequency-Medication, Duration-Medication, Medication-Indication, Medication-ADE,* and *Severity-ADE*. One named entity that is not involved in relations is “other signs and symptoms.” Our annotation guidelines are an extension of the i2b2 annotation guidelines [42] and have been iteratively developed by domain experts. Unlike other clinical corpora that annotate entity relations at the sentence level, we annotated entity relations beyond sentence boundaries. Each EHR note was annotated by at least 2 annotators, and the interannotator agreement of .93 kappa was achieved on our annotations.

The resulting annotated data consisted of 667,061 tokens, 48,803 entity mentions (61.7 per note), and 16,022 entity relations (20.3 per note). The relation distributions in these datasets are reported in the last column of Table 1. *Frequency*, *dosage*, and *indication* are the most frequent relations, whereas *duration* and *adverse* relations are less frequent in the corpus. We split the corpus into 602/95/94 train/develop/test sets.

Figure 1 shows the distribution of relation token distance (the number of tokens between a relation entity mention pair). As shown in Figure 1, most relations occurred within a window of up to 9 tokens. On the other hand, some relations connected entities across multiple sentences. The average relation token distance was 7, and the maximum distance was 769.

To formulate the relation identification task, our goal was to learn a function $f(x)$ that mapped an input entity pair (e_l, e_r) to a relation type $y \in Y$, where Y is the set of all possible relation types including *None*, which in our system denotes the existence of no relation between an entity pair. An entity $e_i \in E$ is any observed entity mention within a document $d \in D$. The input entity pair (e_l, e_r) is sampled from all possible entity pairs E x E within the document and is labeled with a relation type if a true relation holds for it; otherwise, it is labeled *None*. The mention pair and the document within which that pair occurs form a machine learning example x in our task. We implemented and evaluated 3 supervised machine learning approaches as described below, and the experiment workflow is shown in Figure 2.

Table 1. Clinical relation types in our corpus. Entity mentions forming relations are in italics.

Relation	Description	Example	#relations[a]
Dosage	An attribute of a medication: the amount of the medication to be taken	She receives *Albuterol 2 puffs* p.o. q4-6h	2643/336/409
Route	An attribute of a medication: how the medication is administered	She receives *Albuterol* 2 puffs *p.o.* q4-6h	1908/269/332
Frequency	An attribute of a medication: frequency of the administration	She receives *Albuterol* 2 puffs p.o. *q4-6h*	2691/351/451
Duration	An attribute of a medication	The patient was treated with *ampicillin* for *2 weeks*	493/95/110
Indication	A causal relation between a medication and indication: why the drug is taken	He later received *chemotherapy* for his *lung cancer*	2301/264/379
Adverse Event	A causal relation between a medication and an injury: the consequence of a medication	Patient’s death was due to *anaphylactic shock* caused by the intravenously administered *penicillin*	717/134/134
Severity	The attribute of an adverse event	He has *severe diarrhea*	1505/259/241

[a]the number of relations for each type (train/develop/test).

Figure 1. The distribution of relation token distance.

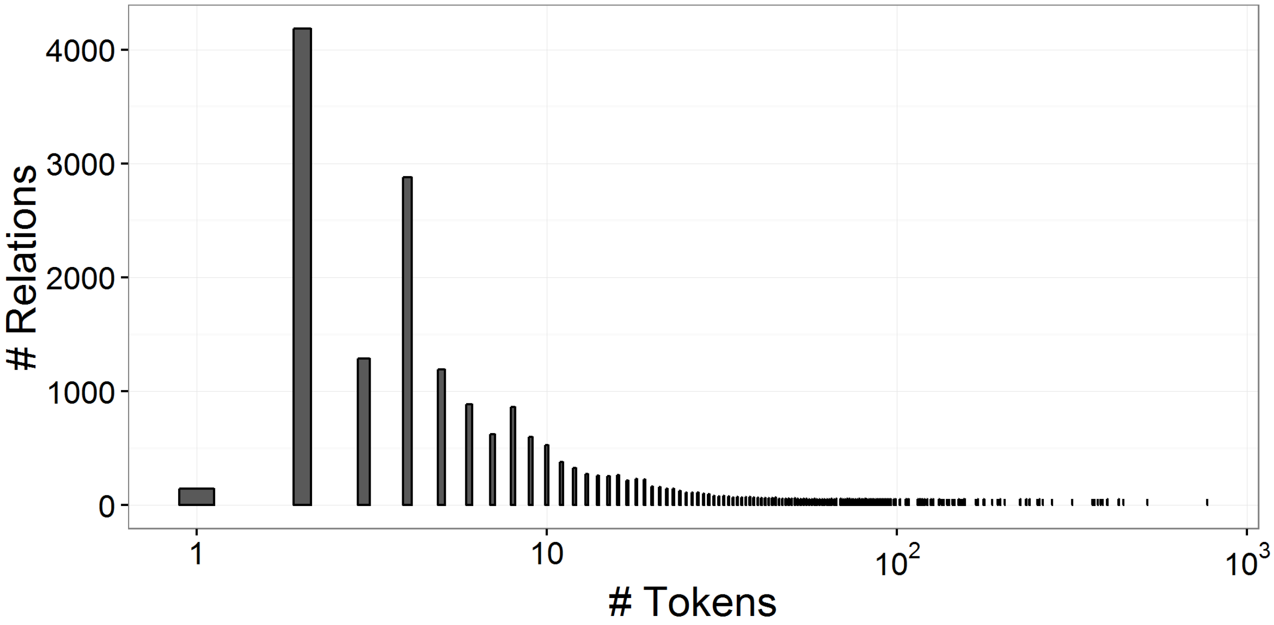
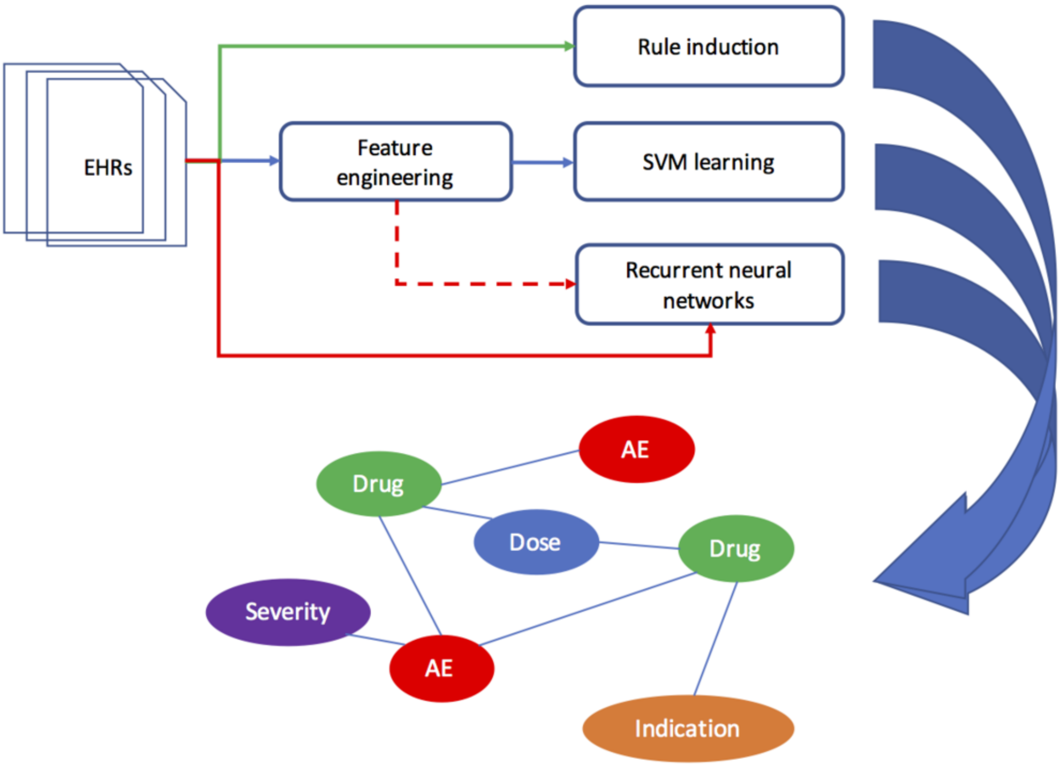


Figure 2. Experimental workflow for adverse drug event (ADE) detection. EHRs: electronic health records; SVM: support vector machines; AE: adverse events.



Induction Rule Baseline

Our first supervised approach used automatically induced rules from the training data, motivated by the observation that the distance between 2 entities was a potentially strong indicator of their relations. For example, we observed that drug attributes typically followed drug names and, in contrast, the distance between adverse drug events and their drugs was relatively far. Therefore, our rule-induction classifier was based on the token distance between 2 entities.

Formally, the classifier considered an entity pair (e_l, e_r) that occurred within a certain distance as a true relation, and the pair was assigned one of the positive relation types, . For training, we calculated the average token distance of the entity pairs for each relation type. We then defined 7 different token distance bins by using these average distances and assigning a single positive relation label to each bin. During prediction, we chose one of the relation labels if the token distance of 2 entities fell in the corresponding bin. For example, if the average token distance for *Severity* relations was 3 and for *Frequency* was 7, we then had 2 bins, $\{ n \mid 0 < n \leq 3\}$ and $\{ n \mid 3 < n \leq 7\}$ (n was the token distance). If the token distance n between an entity pair was in the first bin, the entity pair was given the label *Severity*; otherwise, it was labeled *Frequency* or *None*. We considered an entity pair as *None* relation if their token distance did not belong to any one of the predefined bins.

Support Vector Machines System

We identified a set of rich learning features to build a linear kernel SVM classifier. We chose linear SVM due to its ability to accommodate a large feature space. The features we explored are described below.

Document-level features consisted of the frequencies of a specific entity and entity type in a document.

Relation-specific features were specific to an entity pair being considered for classification. The features were as follows:

- token distance between the 2 entities
- number of clinical entities between the 2 entities
- *n-grams* (*1, 2, 3-grams*) between the 2 entities
- *n-grams* (*1, 2, 3-grams*) of surrounding tokens of the 2 entities. The surrounding tokens were within a window size, which was defined empirically in our experiment.

Entity-level features defined how likely an individual entity mention was involved in a relation:

- one-hot encoding of the left entity type, e_l
- one-hot encoding of the right entity type, e_r
- character *n-grams* (*2, 3-grams*) of the named entities.

Semantic features were derived using the MetaMap tool from National Library of Medicine. Specifically, we mapped entity mentions and their surrounding context to their UMLS(Unified Medical Language System) concepts, preferred terms, and semantic types. We renormalized the concept IDs (identifiers) to their corresponding semantic type names and included both shortened and multiword forms of the semantic types in the feature set. We set the window size of the surrounding context to 10 in the MetaMap tool.

Word representation features were generated to overcome the data sparsity challenge. We explored word clustering and word vector representation features that have been shown to improve performance for chemical and biomedical named-entity recognition tasks [54,55]. In particular, we used the Brown clustering model and Word Vector Classes as word clustering features and applied raw word embedding as word vector features.

We trained the Brown cluster model [56] on a large collection of biomedical text. We then obtained the cluster label prefixes (ie, the top levels of the cluster hierarchy) with 4, 6, 10, and 20 lengths from the Brown model as features for the context of each entity mention. We empirically set the context window size to 10 in this study. To learn broader contextual information, we also explored recently introduced skip-gram model [57]. The skip-gram model is used to predict the contextual words given an input token, and this yielded a dense word embedding for the token that effectively carried its syntactic and semantic information. We first built a skip-gram model on a large unlabeled text consisting of the PubMed abstracts and the EHRs [43], and an additional set of ~2 million PubMed Central full articles. The word embedding induced by the skip-gram model were then clustered into 300 different groups by using a K-means algorithm to obtain cluster labels that we called Word Vector Classes (WVCs). As with the Brown model features, we mapped the entity mention context to their WVCs and included these WVCs in the feature set. We also used the raw word embedding as word representation features in our model, which provided a fine-grained latent feature of word semantic and syntactic information.

The character and word *n-grams* were converted into *TF-IDF(term frequency-inverse document frequency)* weights based on the training set. We stored the *TF-IDF* weights and used them to extract features from the development and test sets. We did not involve the development and test sets in the *n-gram* extraction and the *TF-IDF* calculation to ensure that our models and the features were not biased. We did not extract any sentence-specific features, which allowed us to classify intra- and intersentential relations jointly with a single SVM model.

End-to-End Deep Neural Networks

We explored LSTM and attention-based neural network methods to classify clinical relations in an end-to-end fashion [58] without feature engineering. The reason behind this choice is based on reported advantages of RNNs over CNNs in relation extraction tasks [59,60].

LSTM is a variation of RNN models and was introduced to solve the gradient vanishing problem [61,62]. It can model long-term dependencies with its internal memory, and it achieved notable success with NLP tasks including machine translation [63], speech recognition [64], and textual entailment recognition [65]. The LSTM can effectively learn vector representations for various levels of linguistic units to facilitate different classification tasks. The attention mechanism can help LSTM construct a better representation by selecting important context in an EHR document. As it is computationally expensive to use the whole document for learning the representations, we focused on text windows associated with the 2 entities in our model.

Let x_t, h_t, and c_t be the input, output, and cell state, respectively, at time step t. Given a window of token representations (ie, word embeddings) $x_1,\ldots, x_l$ (x_l is the head token for the entity e_l and L is the window size), an LSTM with hidden size k computes a sequence of the outputs $h_1,\ldots, h_l$ and another sequence of the cell states $c_1,\ldots, c_l$ as: σ

$$i_t = \sigma\,(W_1^{lstm}x_t + W_2^{lstm}h_{t-1} + b_1^{lstm})\ (1)$$

$$i_t' = tanh\,(W_3^{lstm}x_t + W_4^{lstm}h_{t-1} + b_2^{lstm})\ (2)$$

$$f_t = \sigma\,(W_5^{lstm}x_t + W_6^{lstm}h_{t-1} + b_3^{lstm})\ (3)$$

$$o_t = \sigma\,(W_7^{lstm}x_t + W_8^{lstm}h_{t-1} + b_4^{lstm})\ (4)$$

$$c_t = f_t \odot c_{t-1} + i_t \odot i_t'\ (5)$$

$h_t = o_t \odot tanh\,(c_t)$ (6)

where W_1^{lstm},…, $W_8^{lstm} \in R^{k\times k}$ and b_1^{lstm},…, $b_4^{lstm} \in R^k$ are the training parameters, and σ and ⊙ denote the element-wise sigmoid function and the element-wise vector multiplication, respectively.

As described by the equations, the memory cell c_t and hidden state h_t were updated by reading a word token x_t at a time. The memory cell c_t then learns to remember the contextual information that is relevant to the entity mention. This information is then provided to the hidden state h_t by using a gating mechanism, and the last hidden state h_l summarizes all the relevant information for the sequence. i_t, f_t, and o_t are called gates whose values are defined by the nonlinear combination of the previous hidden state h_{t-1} and the current input token x_t and range from 0 to 1. The input gate i_t controls how much information needs to flow into the memory cell, whereas the forget gate f_t decides what information needs to be erased from the memory cell. The output o_t finally produces the hidden state for the current input token.

We further used the output h_l and h_r corresponding to the input token heads of the entity pair e_l and e_r as the entity representations. The representation h_r for entity e_r was obtained similarly by reading its token window with another LSTM. The representations h_l and h_r were then composed by using a function $g\,(h_l, h_r)$ to produce a relation representation r_{lr}. We used a multilayered perceptron (MLP) with a concatenated input for $g\,(h_l, h_r)$ in our model, defined as:

$r_{lr} = g\,(h_l, h_r)$ (7)

$g\,(h_l, h_r) = tanh\,(W_{mlp}[h_l; h_r] + b_{mlp})$ (8)

where $[h_l; h_r]$ is the concatenation operation, $W^{mlp} \in R^{k\times|Y|}$ is the projection matrix, and $b^{mlp} \in R^{|Y|}$ is the bias vector trained from the data. Finally, the relation representation r_{lr} was input to the *softmax* layer to normalize the probability distribution over possible relation types *Y*. The whole network was trained by a backpropagation algorithm by minimizing the cross-entropy loss between the predicted probabilities and the correct labels.

We also experimented LSTM with the attention mechanism, which is expected to solve the issue of the information bottleneck in RNNs [66]. When RNNs process long text, they encounter a practical difficulty; they must compress the text into a single vector with a fixed size. The purpose of the attention mechanism is to exploit the task-relevant outputs in the past time scales and the current output vector to dynamically refine the final vector representation so that the constructed presentation becomes more informative.

We used a standard global attention, which has shown to be state-of-the-art in a variety of NLP tasks: machine translation [66], question answering [67], textual entailment [68], and constituency parsing [69]. In addition to the last output vectors h_l and h_r, the global attention explicitly considered all the previous output vectors h_1,…, h_{l-1} and h_1,…, h_{r-1} to construct attention-weighted representations of the entities e_l and e_r.

Concretely, let $S \in R^{k\times l}$ be a matrix of the output vectors h_1,…, h_l and $o_l \in R^l$ be a vector of ones. An attention weight vector *a*, an attention representation *z*, and the final entity representation $h_l^{'}$ were defined as:

$M = tanh\,(W_1^{at}S + W_2^{at}h_l \oplus o_l)$ (9)

$a = softmax\,(w^T M)$ (10)

$z = Sa^T$ *(11)*

$h_l^{'} = tanh\,(W_3^{at}z + W_4^{at}h_l)$ (12)

where W_1^{at}, W_2^{at}, W_3^{at}, $W_4^{at} \in R^{k\times k}$ are learnable matrices and w^T is the transpose of the learnable vector $w \in R^k$. With the outer product $W_2^{at}h_l \oplus o_l$, we repeated the transformed vector of $h_l l$ times and then combined the resulting matrix with the projected output vectors. The entity representation $h_r^{'}$ for entity e_r was obtained similarly. As for the LSTM-based relation representation, the compositions of the representations were input to an MLP for relation classification.

We also used the bidirectional version of the aforementioned models by feeding concatenated outputs of the forward and backward LSTM. Due to the concatenated outputs, the size of the *W* matrices and *w* vector now become $2\,k \times 2\,k$ and $2\,k$, respectively, increasing the number of parameters to be trained. We have previously shown that bidirectional LSTM outperformed the LSTM models for medication and adverse drug event named-entity recognition tasks in EHRs [43].

Experimental Setup and Evaluation Metrics

As noted previously, we split the corpus into 602/95/94 train/development/test sets. To cast the task as a multiclass classification problem, we generated *None* relations (negative examples) by replacing one of the entity mentions of a true relation with another entity. In doing so, the only constraint was that the new relation should not exist in the true relation corpus set and the rest should be learned from the data. This process gave us additional negative relation instances of 1,190,328/144,338/202,065 for the train/development/test sets, respectively. For this SVM model, we carried out a grid search over its hyperparameters by using the development set for evaluation. Once the best parameters were found, the final SVM model was learned using the optimized hyperparameters on both the training and development sets.

We used ADAM (adaptive moment estimation) [70] for optimization of the neural models. The size of the LSTM hidden units was set to 100. An additional layer was used to map word vectors to the LSTM input. We used a pretrained word2vec model with a size of 300 [43] for word embedding. All neural models were regularized by using 20% input and 30% output dropouts [71] and an l_2 regularizer with strength value 1e-3. The neural models were trained only on the training set. We used the development set to evaluate them for each epoch to choose the best model. The unidirectional models were given 30 epochs and the attentional and bidirectional models were given 60 epochs to converge to an optimum. The final performance of the methods was reported and compared by using the test set.

Our experiment was guided by macro-averaged precision, recall, and *F* 1-score in terms of positive relation types. False negative (*FN*) and false positives (*FP*) are incorrect negative and positive predictions, respectively. True positive (*TP*) results correspond to correct positive predictions, which were actually correct predictions. Recall (*r*) denotes the percentage of correctly labeled positive results over all positive cases and is calculated as: $r=TP/(TP+FN)$. Precision (*p*) is the percentage of correctly labeled positive results over all positive-labeled examples and is calculated as: $p=TP/(TP+FP)$. The *F* 1-measure is the harmonic average of precision and recall, and a balanced *F* 1-score is expressed as: $F_1=2pr/(p+r)$.

Results

This section presents the results of implementing our relation identification systems. We analyzed the performance of each model and the effects of their free parameters.

The Rule Induction Baseline

For this baseline, the distance bins were defined by using the training data. If the token distance of an entity pair did not belong to any of the bins, it was labeled as a *None* relation. This baseline achieved an 7.47% overall F1-score on the test set. Detailed results are shown in Table 2. The performance was low, as the method was very simple. The *Dosage* relation type achieved the highest F1-score (30%) among different relations.

Support Vector Machines–Based Pipeline System

We performed down-sampling for the negative relations (*None* relations) with varying keep rates to study how the performance changed for different distributions of *None* examples involved in the training set. The development and test sets were kept the same.

Table 3 reports the overall F1-score of our SVM model. A higher keep rate means that we used more negative relations in the training set, and that the higher keep rate yielded a better result on the test set in our experiment. We obtained the highest performance with the keep rate value equal to 80% in our SVM model. The training set for this run consisted of 1,096,600 instances, of which 964,520 were *None* relations. In Table 4, we show the detailed performance metrics for this model for each relation type when evaluated on the test set. The F1-scores for most relation types were over 80% with *Route* relation achieving the best of 96%, and the recall of our clinical relation extractor was relatively high. However, the performance of the *Indication* and *Adverse* relations were not as high as those of the other relations, and *Indication* showed the worst score of 75%. We observed that 2 entities forming these types of relations tended to be far away from each other and spanned multiple sentences (the average token distance was 19 and 14, and the maximum was 518 and 769). The long distance makes this relation more difficult to detect than other relations.

Table 2. Results (%) of rule induction classifier on test set.

Relation	Precision	Recall	F1-score
None	100	94	97
Dosage	20	63	30
Route	7	31	11
Frequency	2	7	3
Duration	1	4	1
Indication	1	14	2
Adverse	1	24	1
Severity	0	0	0
Overall	4.57	20.42	7.47

Table 3. Overall F1-scores (%) of support vector machines system. Keep rate for negative down-sampling is varied.

Keep rate	Train	Development	Test
0.1	99.99	99.97	82.46
0.3	99.96	99.93	87.84
0.5	99.94	99.86	89.0
0.8	99.89	99.8	*89.1* [a]

[a]Best score on test data are highlighted in italics.

End-to-End Deep Neural Networks

We also examined the performance of the neural network models. Notably, by leveraging recent advances in deep learning, including efficient representation learning and attention mechanisms, we addressed the problem without any hand-engineered features.

As stated earlier in the Methods section, we used a free parameter window size to determine how much local context is considered for entity representation in neural network models. We first examined the effect of this parameter by training the unidirectional LSTM-based model that was the least complex and the fastest to train and to test. The keep rate for down-sampling was set to 0.1 and the window sizes 5, 10, 30, 50, and 70 were studied. Table 5 presents the results.

When we considered more context with a larger token window, the performance of the LSTM-based relation extractor improved. However, there appeared to be a small drop starting at the point where size is equal to 50, suggesting that large window size may introduce contextual noise into the model. In addition, the training and test time dramatically increased with the large windows; therefore, we set the window size to 30 in our experiments, unless specified.

We conducted a similar group of experiments to observe how the different down-sampling rates affected the model learning. Again, we used an LSTM-based model to report the results, because it was the least complex and fastest to train. The results are presented in Table 6. This time we observed a different pattern of results. The training error kept decreasing as we included more negative examples in the training set. However, with the keep rate of 0.8, it started showing decreasing performance on the development and the test sets. We used a down-sampling keep rate of 0.5 throughout the experiment.

Table 4. Results (%) of the best performing support vector machines model on test set. Keep rate=0.8.

Relation	Precision	Recall	F1-score
None	100	100	100
Dosage	85	91	88
Route	96	97	96
Frequency	93	97	95
Duration	89	93	91
Indication	72	77	75
Adverse	85	84	85
Severity	95	94	95
Overall	87.85	90.42	89.1

Table 5. Overall F1-score of the long short-term memory (LSTM)–based model. Keep rate=0.1.

Window size	Train	Development	Test
5	24.05	14.09	14.58
10	23.92	14.85	14.56
30	37.40	21.77	*22.59* [a]
50	32.1	17.15	18.43
70	27.62	15.04	15.93

[a] Best score on test data are highlighted in italics.

Table 6. Overall F1-score of the long short-term memory (LSTM)–based model. Keep rate for negative down-sampling is varied. Window size=10.

Keep rate	Train	Development	Test
0.1	23.92	14.85	14.56
0.3	38.91	35.18	37.21
0.5	51.25	39.02	*39.45* [a]
0.8	24.82	23.65	21.11

[a] Best score on test data are highlighted in italics.

Table 7. Overall F1-score (%) of long short-term memory (LSTM) and attention-based models. Keep rate=0.5, window size=30.

Model	Train	Development	Test
LSTM[a]	54.47	41.43	42.32
Bidirectional LSTM	86.56	66.47	62.79
LSTM + Attention	68.69	52.71	54.21
Bidirectional LSTM + Attention	83.71	68.95	*65.72* [b]

[a] LSTM: Long short-term memory.

[b] Best score on test data are highlighted in italics.

Table 8. Results (%) of the best-performing neural model (Bidirectional long short-term memory [LSTM] + Attention) on test set. Keep rate=0.5, window size=30.

Relation	Precision	Recall	F1-score
None	100	100	100
Dosage	78	80	79
Route	67	78	72
Frequency	61	76	68
Duration	54	69	61
Indication	32	32	32
Adverse	78	46	58
Severity	77	93	84
Overall	63.85	67.71	65.72

Table 7 shows the performance of variations of the neural models, including the attention-based and the bidirectional LSTM-based relation extractors. The attention-based models always performed better than their corresponding LSTM-based extractors. Furthermore, the bidirectional networks achieve much higher performance than the unidirectional ones. The bidirectional LSTM-based model yielded the highest F-1 training score. However, without the attention mechanism, this model appears to be overfitting. The best performance we obtained on the test set was a 65.72% overall F1-score for positive relation types, which was lower than the one we reported with SVM models. Table 8 shows the detailed test performance measures of the best-performing neural model (bidirectional LSTM + attention) for each relation type. Most of the relation types had F-1 scores above 70%, and *Severity* relation achieved the best performance of 84%. However, the scores for *Indication, Adverse,* and *Duration* relations were relatively low, with the *Indication* score being the lowest of 32%, which is consistent with SVM models. Nevertheless, the overall result is still promising, given the fact that no feature engineering was conducted and that the training set had only hundreds of examples.

For SVM models, we performed an efficient grid search over hyper-parameters, and this boosted performance substantially. However, we were not able to do the same for neural network models due to their computational complexity. Instead, we were able to perform a small random search for neural network parameters.

Discussion

Principal Findings

The bidirectional LSTM model with attention achieved the best performance among all the RNN variations, and additional features are shown to help boost the system performance. SVM model yields the best results, outperforming RNN models, but RNN models demonstrate great potential of significant improvement with more annotated data available.

Both the classic feature engineering-based SVM pipeline and the end-to-end neural network methods have advantages. The SVM model is able to exploit high-dimensional sparse representation (ie, *TF-IDF*), which has traditionally proven to be efficient in clinical NLP tasks. On the other hand, the neural model relies on dense low-dimensional representations that can possibly be constructed in unsupervised fashion from a large unlabeled text, eluding the complicated feature engineering efforts.

However, the neural models have a large number of training parameters that are tuned during training and are able to learn from a much larger dataset for better performance. For example, our bidirectional LSTM model has 1.4 million training parameters, so tuning this parameter set requires a large amount of data. Unfortunately, it is not trivial to obtain such labeled data in the clinical and biomedical domains. Our training data used in the experiments had hundreds of examples per relation type, which was a very small fraction compared with the bidirectional LSTM training parameters. In general, this is a disadvantage of deep learning approaches, and we empirically validated in our ADE relation identification tasks. In low-resource domains, such as the medical domain explored in this study, the focus of future work needs to be on data-efficient deep learning methods. In addition, the SVM relation extractor is easy to train and is robust with a small dataset. Training of the neural network-based relation extractor requires a graphic processing unit (GPU) and is computationally expensive. For example, 60 epochs of our attention model took 26 hours to complete on a GeForce GTX 980 GPU.

Error Analysis

We analyzed how well the SVM and attention models performed on short- and long-distance relations. Figure 3 plots the test F1-score of these models against relation distance. The bidirectional LSTM with attention did not perform well on short distance relations, and it was not stable. In contrast, SVM was very stable and performed well for those relations where the distances between the entities are long. Interestingly, the neural network performance decreased to 87% from 100% when the distance was 1100. The performance drop was due to false positives, and the generated negative examples were classified as positive by the model. However, these were the simple cases that even our rule induction classifier was able to easily detect. Therefore, we hypothesize that the neural network makes this obvious mistake because the context features, such as relation representations the model relies on, are not sufficient for the task. To justify this, we included a set of additional features in the neural network model. The token and mention distances and mention type features (in SVM models) were embedded and further used along with the dense-vector relation representations for classification.

By including these additional features in the neural model, we improved its best result from a 65.72% to a 77.35% F1-score. Table 9 provides a horizontal comparison of the different methods proposed in this paper. Inclusion of those features in the neural model yielded an approximately 12% improvement, and the performance gap between the neural model and SVM model was also reduced.

We also conducted a set of experiments to show how the training data size affects the overall performance of the SVM and neural models. We created new training sets with stratified sampling rates of 20%, 40%, 60%, and 80% of the original training data. Both SVM and attention-based bidirectional LSTM models were trained on the new training sets and evaluated on the test data. In Figure 4, we display the test F1-scores of the models for different sample sizes. The SVM model achieved an F1-score greater than 80% even when trained on 20% of the data, but the performance of the neural model was only around 62%. This demonstrates that feature engineering approach may be preferred over deep learning models when less annotated data are available, as the hand-crafted features in the SVM model has encoded human knowledge, such as domain knowledge and various heuristics.

However, as the training dataset is increased, we can observe a firm improvement on the performances of the neural models. When we increased the training sample size from 20% to 80%, the neural model improved the test performance from ~62% to ~76, by almost 20%, whereas the improvement range for the SVM model was much smaller, around 8% F1-score. Therefore, the neural model has the potential to improve substantially if a larger training dataset is available.

Limitations

One limitation of this study is that the size of the data in the experiment is relatively small, and more follow-up study is needed to further verify the findings on a larger dataset or other publicly available datasets (eg, i2b2 data although they only contain intrasentential relations) by exploring more RNN or CNN architectures, which we will investigate in our future work. In addition, the global attention in our LSTM model may not be sufficient to pinpoint important local context, especially for long-distance relations, and it is worth exploring more flexible attention mechanisms on this task.

Figure 3. Test F1-score over relation distance. BiLSTM: bidirectional long short term memory; SVM: support vector machine.

Figure 4. Test F1-score over varying training sample size. BiLSTM: bidirectional long short term memory; SVM: support vector machine.

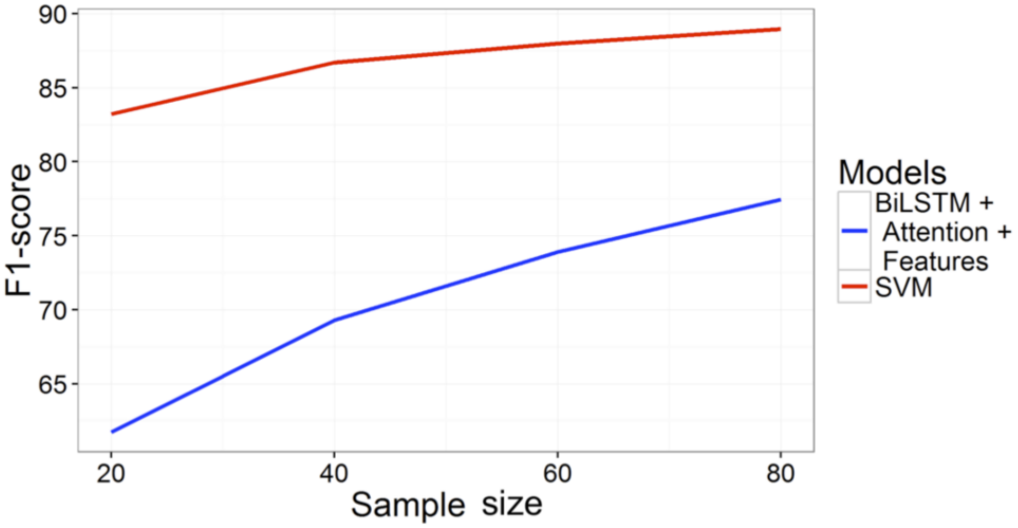


Table 9. Comparison of different models in terms of overall F1-score.

Model	Train	Development	Test
Rule induction classifier	8.33	8.74	7.47
Bidirectional LSTM[b]	83.71	66.47	62.79
Bidirectional LSTM + Attention	86.56	68.95	65.72
Bidirectional LSTM + Attention + Features	88.14	77.77	77.35
SVM[a] + Features	87.85	90.42	*89.1* [c]

[a] LSTM: Long short-term memory

[b] SVM: support vector machines.

[c] Best score on test data are highlighted in italics.

Conclusions

In this study, we created a new expert-annotated EHR corpus in the context of ADE relation identification, which will become a valuable resource and benchmark in drug safety surveillance research community. We, then, explored 3 different supervised machine learning models with different levels of complexity to identify 7 types of ADE-related clinical relations. Our results show that the SVM model with a rich feature set achieved the highest performance, surpassing both the rule induction model and the RNN models. The bidirectional LSTM model with attention achieved the best performance among the RNN models, and the additional features are shown to help boost the system performance. However, its performance remains substantially inferior to the performance of the SVM model, although RNN models demonstrate great potential of significant improvement with more annotated data available. Our results indicate that a rich feature set remains crucial for relation identification in clinical text, especially when the training size is small.

In the future, we will further explore different deep learning architectures (eg, multikernel CNNs, hierarchical RNNs, multilevel attentions) on this task for improved performance. Then, we plan to apply our system to EHRs on a large scale and derive meaningful insights to facilitate efficient and effective drug safety surveillance.

Acknowledgments

This work was supported by the grant R01HL125089 from the National Institutes of Health. Any opinions, findings, and conclusions or recommendations expressed in this paper are those of the authors and do not necessarily reflect those of the sponsor.

Conflicts of Interest

None declared.

References

1. Haas JS, Iyer A, Orav EJ, Schiff GD, Bates DW. Participation in an ambulatory e-pharmacovigilance system. Pharmacoepidemiol Drug Saf 2010 Sep;19(9):961-969. [doi: 10.1002/pds.2006] [Medline: 20623512]

2. Bates DW, Cullen DJ, Laird N, Petersen LA, Small SD, Servi D, et al. Incidence of adverse drug events and potential adverse drug events. Implications for prevention. ADE Prevention Study Group. J Am Med Assoc 1995 Jul 05;274(1):29-34. [Medline: 7791255]
3. Classen D, Pestonik S, Scott ER, Lloyd J, Burke J. Adverse drug events in hospitalized patientsxcess length of stay, extra costs, and attributable mortality. J Am Med Assoc 1997;277(4):e301-e306. [Medline: 9002492]
4. Bates DW, Spell N, Cullen DJ, Burdick E, Laird N, Petersen LA, et al. The costs of adverse drug events in hospitalized patients. Adverse Drug Events Prevention Study Group. J Am Med Assoc 1997;277(4):307-311. [Medline: 9002493]
5. Nebeker JR, Hoffman JM, Weir CR, Bennett CL, Hurdle JF. High rates of adverse drug events in a highly computerized hospital. Arch Intern Med 2005 May 23;165(10):1111-1116. [doi: 10.1001/archinte.165.10.1111] [Medline: 15911723]
6. Handler SM, Altman RL, Perera S, Hanlon JT, Studenski SA, Bost JE, et al. A systematic review of the performance characteristics of clinical event monitor signals used to detect adverse drug events in the hospital setting. J Am Med Inform Assoc 2007 Jul;14(4):451-458 [FREE Full text] [doi: 10.1197/jamia.M2369] [Medline: 17460130]
7. Lazarou J, Pomeranz B, Corey P. Incidence of adverse drug reactions in hospitalized patients: a meta-analysis of prospective studies. J Am Med Assoc 1998;279(15):1200-1205. [Medline: 9555760]
8. Classen D, Pestotnik S, Evans R, Burke J. Description of a computerized adverse drug event monitor using a hospital information system. Hosp Pharm 1992;27(9):783. [Medline: 10121426]
9. Kaushal R, Jha AK, Franz C, Glaser J, Shetty KD, Jaggi T, Brigham and Women's Hospital CPOE Working Group. Return on investment for a computerized physician order entry system. J Am Med Inform Assoc 2006 May;13(3):261-266 [FREE Full text] [doi: 10.1197/jamia.M1984] [Medline: 16501178]
10. World Health Organization (WHO). Pharmacovigilance URL: http://www.who.int/medicines/areas/quality_safety/safety_efficacy/pharmvigi/en/ [WebCite Cache ID 6uhg58vQS]
11. Edlavitch SA. Adverse drug event reporting. Improving the low US reporting rates. Arch Intern Med 1988 Jul;148(7):1499-1503. [Medline: 3382293]
12. Rogers AS, Israel E, Smith CR, Levine D, McBean AM, Valente C, et al. Physician knowledge, attitudes, and behavior related to reporting adverse drug events. Arch Intern Med 1988 Jul;148(7):1596-1600. [Medline: 3382304]
13. Begaud B, Moride Y, Tubert-Bitter P, Chaslerie A, Haramburu F. False-positives in spontaneous reporting: should we worry about them? Br J Clin Pharmacol 2012 Jul 05;38(5):401-404. [doi: 10.1111/j.1365-2125.1994.tb04373.x]
14. Xu R, Wang Q. Comparing a knowledge-driven approach to a supervised machine learning approach in large-scale extraction of drug-side effect relationships from free-text biomedical literature. BMC Bioinformatics 2015;16 Suppl 5:S6 [FREE Full text] [doi: 10.1186/1471-2105-16-S5-S6] [Medline: 25860223]
15. Butt TF, Cox AR, Oyebode JR, Ferner RE. Internet accounts of serious adverse drug reactions: a study of experiences of Stevens-Johnson syndrome and toxic epidermal necrolysis. Drug Saf 2012 Dec 01;35(12):1159-1170. [doi: 10.2165/11631950-000000000-00000] [Medline: 23058037]
16. CISION. 2013. Adverse event reporting: What pharmaceutical companies need to know URL: http://www.cision.com/us/2013/12/adverse-event-reporting-pharma/ [WebCite Cache ID 6uhyRoqPe]
17. Lardon J, Abdellaoui R, Bellet F, Asfari H, Souvignet J, Texier N, et al. Adverse drug reaction identification and extraction in social media: a scoping review. J Med Internet Res 2015 Jul 10;17(7):e171 [FREE Full text] [doi: 10.2196/jmir.4304] [Medline: 26163365]
18. Abdellaoui R, Schück S, Texier N, Burgun A. Filtering entities to optimize identification of adverse drug reaction from social media: how can the number of words between entities in the messages help? JMIR Public Health Surveill 2017 Jun 22;3(2):e36 [FREE Full text] [doi: 10.2196/publichealth.6577] [Medline: 28642212]
19. Rossi AC, Knapp DE, Anello C, O'Neill RT, Graham CF, Mendelis PS, et al. Discovery of adverse drug reactions. J Am Med Assoc 1983 Apr 22;249(16):2226. [doi: 10.1001/jama.1983.03330400072029]
20. Gurwitz J, Field T, Harrold L, Rothschild J, Debellis K, Seger A, et al. Incidence and preventability of adverse drug events among older persons in the ambulatory setting. J Am Med Assoc 2003;289(9):1107-1116. [Medline: 12622580]
21. FDA. Questions and Answers on FDA's Adverse Event Reporting System (FAERS) URL: http://www.fda.gov/Drugs/GuidanceComplianceRegulatoryInformation/Surveillance/AdverseDrugEffects/default.htm [accessed 2018-03-04] [WebCite Cache ID 6uhyyje6x]
22. McGraw D, Rosati K, Evans B. A policy framework for public health uses of electronic health data. Pharmacoepidemiol Drug Saf 2012 Jan;21(Suppl 1):18-22. [doi: 10.1002/pds.2319] [Medline: 22262589]
23. Honigman B, Lee J, Rothschild J, Light P, Pulling R, Yu T, et al. Using computerized data to identify adverse drug events in outpatients. J Am Med Inform Assoc 2001;8(3):254-266. [Medline: 11320070]
24. Brown JS, Kulldorff M, Petronis KR, Reynolds R, Chan KA, Davis RL, et al. Early adverse drug event signal detection within population-based health networks using sequential methods: key methodologic considerations. Pharmacoepidemiol Drug Saf 2009;18(3):226-234. [Medline: 19148879]
25. Liu M, McPeek HE, Matheny ME, Denny JC, Schildcrout JS, Miller RA, et al. Comparative analysis of pharmacovigilance methods in the detection of adverse drug reactions using electronic medical records. J Am Med Inform Assoc 2013 May 01;20(3):420-426 [FREE Full text] [doi: 10.1136/amiajnl-2012-001119] [Medline: 23161894]

26. Hurdle JF, Weir CR, Roth B, Hoffman J, Nebeker JR. Critical gaps in the world's largest electronic medical record: Ad Hoc nursing narratives and invisible adverse drug events. AMIA Annu Symp Proc 2003:309-312 [FREE Full text] [Medline: 14728184]
27. Gurulingappa H, Mateen-Rajput A, Toldo L. pdfs.semanticscholar. Extraction of potential adverse drug events from medical case reports URL: https://pdfs.semanticscholar.org/8352/a732f635b6071026d165cb920e6e5d0cc934.pdf [accessed 2018-03-16] [WebCite Cache ID 6xybziLU3]
28. Kang N, Singh B, Bui C, Afzal Z, van Mulligen EM, Kors JA. Knowledge-based extraction of adverse drug events from biomedical text. BMC Bioinformatics 2014 Mar 04;15:64 [FREE Full text] [doi: 10.1186/1471-2105-15-64] [Medline: 24593054]
29. Leaman R, Wojtulewicz L, Sullivan R, Skariah A, Yang J, Gonzalez G. Towards Internet-Age Pharmacovigilancextracting Adverse Drug Reactions from User Posts to Health-Related Social Networks. In: Proceedings of the 2010 Workshop on Biomedical Natural Language Processing. 2010 Presented at: BioNLP '10; July 15-15, 2010; Uppsala, Sweden p. 117-125.
30. Polepalli RB, Belknap SM, Li Z, Frid N, West DP, Yu H. Automatically recognizing medication and adverse event information from food and drug administration's adverse event reporting system narratives. JMIR Med Inform 2014 Jun 27;2(1):e10 [FREE Full text] [doi: 10.2196/medinform.3022] [Medline: 25600332]
31. Visweswaran S, Hanbury P, Saul M, Cooper GF. Detecting adverse drug events in discharge summaries using variations on the simple Bayes model. AMIA Annu Symp Proc 2003:689-693 [FREE Full text] [Medline: 14728261]
32. Phansalkar S, South BR, Hoffman JM, Hurdle JF. Looking for a needle in the haystack? A case for detecting adverse drug events (ADE) in clinical notes. AMIA Annu Symp Proc 2007 Oct 11:1077. [Medline: 18694175]
33. Iqbal E, Mallah R, Jackson RG, Ball M, Ibrahim ZM, Broadbent M, et al. Identification of adverse drug events from free text electronic patient records and information in a large mental health case register. PLoS One 2015 Aug;10(8):e0134208 [FREE Full text] [doi: 10.1371/journal.pone.0134208] [Medline: 26273830]
34. Aramaki E, Miura Y, Tonoike M, Ohkuma T, Masuichi H, Waki K, et al. Extraction of adverse drug effects from clinical records. Stud Health Technol Inform 2010;160:739-743. [Medline: 20841784]
35. Henriksson A, Kvist M, Dalianis H, Duneld M. Identifying adverse drug event information in clinical notes with distributional semantic representations of context. J Biomed Inform 2015 Aug 17;57:333-349 [FREE Full text] [doi: 10.1016/j.jbi.2015.08.013] [Medline: 26291578]
36. Casillas A, Pérez A, Oronoz M, Gojenola K, Santiso S. Learning to extract adverse drug reaction events from electronic health records in Spanish. Expert Syst Appl 2016 Nov;61:235-245. [doi: 10.1016/j.eswa.2016.05.034]
37. Wang G, Jung K, Winnenburg R, Shah NH. A method for systematic discovery of adverse drug events from clinical notes. J Am Med Inform Assoc 2015 Nov;22(6):1196-1204 [FREE Full text] [doi: 10.1093/jamia/ocv102] [Medline: 26232442]
38. LePendu P, Iyer SV, Bauer-Mehren A, Harpaz R, Mortensen JM, Podchiyska T, et al. Pharmacovigilance using clinical notes. Clin Pharmacol Ther 2013 Jun;93(6):547-555 [FREE Full text] [doi: 10.1038/clpt.2013.47] [Medline: 23571773]
39. Personeni G, Bresso E, Devignes M, Dumontier M, Smaïl-Tabbone M, Coulet A. Discovering associations between adverse drug events using pattern structures and ontologies. J Biomed Semantics 2017 Aug 22;8(1):29 [FREE Full text] [doi: 10.1186/s13326-017-0137-x] [Medline: 28830518]
40. Banda J, Evans L, Vanguri R, Tatonetti N, Ryan P, Shah N. A curated and standardized adverse drug event resource to accelerate drug safety research. Sci Data 2016 Dec 10;3:160026 [FREE Full text] [doi: 10.1038/sdata.2016.26] [Medline: 27193236]
41. Uzuner O, Solti I, Cadag E. Extracting medication information from clinical text. J Am Med Inform Assoc 2010;17(5):514-518 [FREE Full text] [doi: 10.1136/jamia.2010.003947] [Medline: 20819854]
42. Uzuner Ö, South B, Shen S, DuVall S. 2010 i2b2/VA challenge on concepts, assertions, and relations in clinical text. J Am Med Inform Assoc 2011;18(5):552-556. [doi: 10.1136/amiajnl-2011-000203]
43. Jagannatha AN, Yu H. Bidirectional RNN for medical event detection in electronic health records. Proc Conf 2016 Jun;2016:473-482 [FREE Full text] [Medline: 27885364]
44. Collobert R, Weston J, Bottou L, Karlen M, Kavukcuoglu K, Kuksa P. Natural language processing (almost) from scratch. J Mach Learn Res 2011;12:2493-2537.
45. Andor D, Alberti C, Weiss D, Severyn A, Presta A, Ganchev K, et al. Globally Normalized Transition-Based Neural Networks. 2016 Presented at: the 54th Annual Meeting of the Association for Computational Linguistics; August 7-12; Berlin, Germany p. 2442-2452.
46. Yan X, Mou L, Li G, Chen Y, Peng H, Jin Z. Classifying Relations via Long Short Term Memory Networks along Shortest Dependency Path. 2015 Presented at: Conf Empir Methods Nat Lang Process; September 17-21; Lisbon, Portugal p. 1785-1794.
47. Peng N, Poon H, Quirk C, Toutanova K, Yih W. Cs.jhu.edu. 2017. Cross-Sentence N-ary Relation Extraction with Graph LSTMs URL: https://www.cs.jhu.edu/~npeng/papers/TACL_17_RelationExtraction.pdf [accessed 2018-03-17] [WebCite Cache ID 6xychvITD]
48. Wang L, Cao Z, Melo GD, Liu Z. Relation Classification via Multi-Level Attention CNNs. 2016 Presented at: the 54th Annual Meeting of the Association for Computational Linguistics; August 7-12; Berlin, Germany p. 1298-1307.

49. Miwa M, Bansal M. End-to-end Relation Extraction using LSTMs on Sequences and Tree Structures. 2016 Presented at: Proc ACL; August 7-12; Berlin, Germany.
50. Li F, Zhang M, Fu G, Ji D. A neural joint model for entity and relation extraction from biomedical text. BMC Bioinformatics 2017 Mar 31;18(1):198 [FREE Full text] [doi: 10.1186/s12859-017-1609-9] [Medline: 28359255]
51. Lv X, Guan Y, Yang J, Wu J. Clinical Relation Extraction with Deep Learning. IJHIT 2016 Jul 31;9(7):237-248. [doi: 10.14257/ijhit.2016.9.7.22]
52. Sahu S, Anand A, Oruganty K, Gattu M. arxiv.org. 2016. Relation extraction from clinical texts using domain invariant convolutional neural network URL: https://arxiv.org/pdf/1606.09370.pdf [accessed 2018-03-17] [WebCite Cache ID 6xycvyTGd]
53. Novak P, Lavrač N, Webb G. Supervised descriptive rule induction. In: Encyclopedia of Machine Learning. Boston, MA: Springer; 2011.
54. Munkhdalai T, Li M, Batsuren K, Park HA, Choi NH, Ryu KH. Incorporating domain knowledge in chemical and biomedical named entity recognition with word representations. J Cheminform 2015;7(Suppl 1 Text mining for chemistry and the CHEMDNER track):S9 [FREE Full text] [doi: 10.1186/1758-2946-7-S1-S9] [Medline: 25810780]
55. Zheng J, Yarzebski J, Ramesh B, Goldberg R, Yu H. Automatically detecting acute myocardial infarction events from EHR text: a preliminary study. AMIA Annu Symp Proc 2014;2014:1286-1293 [FREE Full text] [Medline: 25954440]
56. Brown PF, Desouza PV, Mercer RL, Pietra VJ, Lai JC. Class-based n-gram models of natural language. Comput Linguist 1992;18(4):479.
57. Mikolov T, Chen K, Corrado G, Dean J. arxiv.org. 2013. Efficient estimation of word representations in vector space URL: https://arxiv.org/pdf/1301.3781.pdf [accessed 2018-03-17] [WebCite Cache ID 6xydGYKsS]
58. Glasmachers T. proceedings.mlr.press. 2017. Limits of End-to-End Learning URL: http://proceedings.mlr.press/v77/glasmachers17a/glasmachers17a.pdf [accessed 2018-03-17] [WebCite Cache ID 6xydP34zO]
59. Zhang D, Wang D. Relation Classification: CNN or RNN? In: Natural Language Understanding and Intelligent Applications. Cham: Springer; 2016:665-675.
60. Zhang D, Wang D. arXiv.org. 2015. Relation Classification via Recurrent Neural Network URL: https://arxiv.org/pdf/1508.01006.pdf [accessed 2018-03-17] [WebCite Cache ID 6xydY8IOW]
61. Bengio Y, Simard P, Frasconi P. Learning long-term dependencies with gradient descent is difficult. Neural Netw IEEE Trans 1994;5(2):157-166. [doi: 10.1109/72.279181]
62. Hochreiter S. The vanishing gradient problem during learning recurrent neural nets and problem solutions. Int J Uncertain Fuzziness Knowl-Based Syst 1998;6(2):107. [doi: 10.1142/S0218488598000094]
63. Sutskever I, Vinyals O, Le QV. Sequence to Sequence Learning with Neural Networks. 2014 Presented at: NIPS; December 08 - 13; Montreal, Canada p. 3104-3112.
64. Graves A, Mohamed A, Hinton G. Speech Recognition with Deep Recurrent Neural Networks. 2013 Presented at: IEEE ICASSP; May 26-31; Vancouver, BC, Canada p. 6645-6649. [doi: 10.1109/ICASSP.2013.6638947]
65. Bowman SR, Angeli G, Potts C, Manning CD. A large annotated corpus for learning natural language inference. In: Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing. 2015 Presented at: EMNLP; September 17-21; Lisbon, Portugal.
66. Boehning D, Cho K, Bengio Y. arxiv.org. 2015. Neural Machine Translation by Jointly Learning to Align and Translate URL: https://arxiv.org/pdf/1409.0473.pdf [accessed 2018-03-17] [WebCite Cache ID 6xydutE0m]
67. Hermann K, Kočiský T, Grefenstette E, Espeholt L, Kay W, Suleyman M, et al. arXiv.org. 2015. Teaching Machines to Read and Comprehend URL: https://arxiv.org/pdf/1506.03340.pdf [accessed 2018-03-17] [WebCite Cache ID 6xye7stpJ]
68. Rocktäschel T, Grefenstette E, Hermann K, Kočiský T, Blunsom P. arxiv.org. 2015. Reasoning about Entailment with Neural Attention URL: https://arxiv.org/pdf/1509.06664.pdf [accessed 2018-03-17] [WebCite Cache ID 6xyeAy6dt]
69. Vinyals O, Kaiser L, Koo T, Petrov S, Sutskever I, Hinton G. Grammar as a Foreign Language. 2015 Presented at: NIPS; Dec 7-12; Montreal, Canada.
70. Kingma D, Ba J. Adam: a Method for Stochastic Optimization. 2014 Presented at: Int Conf Learn Represent; April 14-16; Banff, Canada p. 1-13.
71. Srivastava N, Hinton G, Krizhevsky A, Sutskever I, Salakhutdinov R. A simple way to prevent neural networks from overfitting. J Mach Learn Res 2014;15(1):1929-1958.

Abbreviations

ADE: adverse drug event
CNNs: convolutional neural networks
EHR: electronic health record
FDA: Food And Drug Administration
FN: false negative
FP: false positives
GPU: graphic processing unit

HER: electronic health record
LSTM: long short-term memory
MLP: multilayered perceptron
NLP: natural language processing
RNN: recurrent neural network
SVM: support vector machines
TP: true positive
WVCs: Word Vector Classes

Edited by G Eysenbach; submitted 08.11.17; peer-reviewed by M Torii, G Gonzalez, M Liu; comments to author 09.12.17; revised version received 03.02.18; accepted 05.02.18; published 25.04.18.

Please cite as:
Munkhdalai T, Liu F, Yu H
Clinical Relation Extraction Toward Drug Safety Surveillance Using Electronic Health Record Narratives: Classical Learning Versus Deep Learning
JMIR Public Health Surveill 2018;4(2):e29
URL: http://publichealth.jmir.org/2018/2/e29/
doi:10.2196/publichealth.9361
PMID:29695376

Proposal

Causality Patterns for Detecting Adverse Drug Reactions From Social Media: Text Mining Approach

Danushka Bollegala[1], PhD; Simon Maskell[1*], PhD; Richard Sloane[1], PhD; Joanna Hajne[1], PhD; Munir Pirmohamed[2*], PhD

[1]Department of Computer Science, University of Liverpool, Liverpool, United Kingdom
[2]Department of Translational Medicine, University of Liverpool, Liverpool, United Kingdom
[*]these authors contributed equally

Corresponding Author:
Danushka Bollegala, PhD
Department of Computer Science
University of Liverpool
Ashton Street,
Liverpool, L693BX
United Kingdom
Phone: 44 1517954283
Fax: 44 1517954283
Email: danushka.bollegala@liverpool.ac.uk

Abstract

Background: Detecting adverse drug reactions (ADRs) is an important task that has direct implications for the use of that drug. If we can detect previously unknown ADRs as quickly as possible, then this information can be provided to the regulators, pharmaceutical companies, and health care organizations, thereby potentially reducing drug-related morbidity and saving lives of many patients. A promising approach for detecting ADRs is to use social media platforms such as Twitter and Facebook. A high level of correlation between a drug name and an event may be an indication of a potential adverse reaction associated with that drug. Although numerous association measures have been proposed by the signal detection community for identifying ADRs, these measures are limited in that they detect correlations but often ignore causality.

Objective: This study aimed to propose a causality measure that can detect an adverse reaction that is caused by a drug rather than merely being a correlated signal.

Methods: To the best of our knowledge, this was the first causality-sensitive approach for detecting ADRs from social media. Specifically, the relationship between a drug and an event was represented using a set of automatically extracted lexical patterns. We then learned the weights for the extracted lexical patterns that indicate their reliability for expressing an adverse reaction of a given drug.

Results: Our proposed method obtains an ADR detection accuracy of 74% on a large-scale manually annotated dataset of tweets, covering a standard set of drugs and adverse reactions.

Conclusions: By using lexical patterns, we can accurately detect the causality between drugs and adverse reaction–related events.

(JMIR Public Health Surveill 2018;4(2):e51) doi:10.2196/publichealth.8214

KEYWORDS

machine learning; ADR detection; causality; lexical patterns; causality detection; support vector machines

Introduction

Background

An adverse drug reaction (ADR) is defined as “an appreciably harmful or unpleasant reaction, resulting from an intervention related to the use of a medicinal product, which predicts hazard from future administration and warrants prevention or specific treatment, alternation of the dosage regimen, or withdrawal of the product” [1-4]. It is estimated that approximately 2 million patients in the United States are affected each year by serious ADRs, resulting in roughly 100,000 fatalities [5]. In fact, ADRs are the fourth leading cause of death in the United States, following cancer and heart diseases [6]. Treating patients who develop ADRs results in significant health costs to nations throughout the world. For example, it has been estimated that US $136 billion is spent each year on treatments related to ADRs in the United States [7,8].

In an ideal world, all adverse reactions associated with a drug need to be detected before marketing, and the drug label modified accordingly. However, this is not feasible due to several reasons. First, the number of human subjects participating in a clinical trial of a premarketed drug is often small, which limits the statistical power to detect ADRs, particularly those which may be uncommon. In fact, rare ADRs are usually not detected during the premarketing phases of drug development. Second, as many of the clinical trials are short-lasting, ADRs which are delayed will not be detected. Third, some ADRs show up only when a drug is being taken together with other drugs, leading to an adverse drug-drug interaction. Considering that the number of combinations of drugs is potentially large, it is impractical to test for all of the possible combinations during a clinical trial. Fourth, *drug repurposing* [9]—the practice of off-label usage of drugs for treating diseases for which they were not originally intended—could lead to unforeseen ADRs.

Due to these challenges in detecting ADRs during the premarketing phase, identification of ADRs in the postmarketing phase remains hugely important. The cornerstone of postmarketing pharmacovigilance remains the spontaneous reporting schemes such as the Yellow Card Scheme [10] in the United Kingdom and the MedWatch system [11] in the United States. Such schemes allow hospitals, medical practitioners, and patients to report ADRs. Unfortunately, the reporting rates are generally poor. For example, only 10% of serious ADRs and 2% to 4% of nonserious ADRs are reported [12].

Although patients experience ADRs, they may be reluctant to report their experiences through official reporting systems for various reasons. For example, patients might be unfamiliar with or unaware of the ADR reporting schemes, or might find it difficult to understand the terminology used in the forms, or might not be aware of the importance of reporting ADRs. Even when ADRs have been reported via such spontaneous reporting systems, the time required from the first report to any regulatory action may be long, which is problematical in protecting public health from iatrogenic conditions.

An alternative approach for detecting ADRs in a timely manner on a larger scale is to use social media. Social media platforms such as Twitter [13], Facebook [14], Instagram [15], and Pinterest [16] have been used extensively for market analysis of various products. Social media provides a convenient and direct access to consumers' opinions about the products and services they use. In comparison with a clinical study, which inevitably is limited to a small number of participants, in social media we can access comments from a massive number of diverse groups of people. Due to its potential value, the pharmacovigilance community has already started to exploit social media as a potential reporting tool for obtaining information about ADRs [17]. For example, the WEB-RADR [18] project funded by the Innovative Medicines Initiative was funded to evaluate the usefulness of social media as a reporting tool for ADRs.

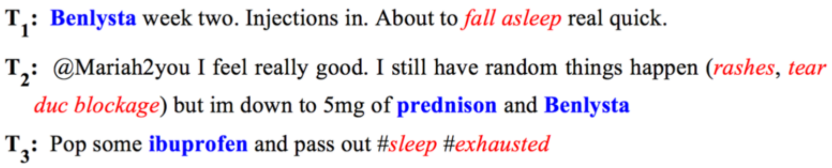
However, compared with spontaneous reporting systems where patients or health care practitioners explicitly report ADRs, detecting ADRs from social media poses several challenges. Because social media is not perceived by most patients as an official reporting tool for ADRs, a drug and its associated ADRs might not be completely expressed in a single social media post. This issue is further aggravated by the limitations imposed on the length of a post in social media platforms. For example, in Twitter, a single post (aka a *tweet*) is limited to a maximum of 140 characters. Even in social media platforms where such limitations do not exist, such as Facebook, the users might not always provide comprehensive reports containing all the information that would normally be completed on a Yellow Card. Furthermore, social media users often interact with social media platforms through specialized apps on mobile devices such as mobile phones, which do not possess physical keyboards that facilitate the entering of longer texts.

In addition to the brevity and incompleteness of social media posts as a medium for reporting ADRs, the reliability of the information expressed through social media is also a concern. It is often difficult to authenticate the information disseminated through social media. For example, in Twitter, the same user can create multiple accounts under different names including aliases. False information might be expressed intentionally or unintentionally in social media, which makes it difficult to verify the information extracted from social media. Unlike in the Yellow Card system, where it is possible to contact a reporter to obtain further information, in social media it is difficult to obtain additional information from users due to anonymity and privacy settings. All of these challenges introduce various levels of noise to ADR signal that can be captured from social media. Consequently, methods that detect ADRs from social media need to overcome these challenges.

An approach for detecting significant signals indicating adverse reactions to drugs in social media is to measure the correlation between a drug and an event. If many social media posts or users mention a drug and an event, then the likelihood that the drug causes an adverse reaction increases. Indeed, numerous measures have been proposed in previous work to measure the degree of association between a drug and an adverse reaction [19-26]. Although co-occurrence measures do not completely solve all of the above-mentioned challenges of using social media, they provide a practical and a highly scalable mechanism for detecting ADRs from social media.

A fundamental drawback of co-occurrence-based approaches for detecting ADRs is that they ignore the context in which a drug and an ADR co-occur in social media. Co-occurrence *does not* always indicate causality. Although a drug and an event that could suggest an ADR might be mentioned frequently in social media, the co-occurrence may be because the drug is used as a remedy for that symptom. Moreover, the drug may have been taken by 1 person, but the social media post mentions the ADR in a different person. However, the context in which a drug and an ADR co-occur can provide useful clues that can be used to separate causality from co-occurrence.

Figure 1. Three tweets mentioning a drug (shown in blue boldface fonts) and symptoms (shown in red italic font).



T_1: **Benlysta** week two. Injections in. About to *fall asleep* real quick.

T_2: @Mariah2you I feel really good. I still have random things happen (*rashes*, *tear duc blockage*) but im down to 5mg of **prednison** and **Benlysta**

T_3: Pop some **ibuprofen** and pass out #*sleep* #*exhausted*

To illustrate the usefulness of contextual information for ADR detection, consider the 3 tweets shown in Figure 1. T_1 is suggestive of an association with a drug and a potential adverse reaction. T_2 may reflect that the patient's disease improving or that an ADR occurred but is waning following dose reduction. T_3 is unlikely to be an ADR; Ibuprofen is being taken by this patient to potentially relieve the pain and have some sleep. These examples show that there are useful hints we can extract from the tweets such as *about to (feel an ADR), I still have (ADRs)* that we can use to evaluate the causality relationship between a mentioned drug and an adverse reaction.

Why is solving this problem critical for systems that attempt to extract ADRs from social media? The standard practice in the pharmacovigilance community for detecting ADRs from patient reports is to apply disproportionality measures that consider *only* co-occurrence (and occurrence) counts. Unfortunately, disproportionality measures by design are agnostic to the linguistic context in social media and are therefore unable to utilize the clues that appear in social media to determine whether an ADR is truly caused by the drug. However, given a tweet containing a drug and a potential adverse reaction, if we can first develop a classifier that predicts whether this tweet is describing a causality relationship, we then can use disproportionality measures on the tweets that are identified as positive by the classifier for further analysis. This preprocessing step is likely to improve the accuracy of the ADR detection process. Moreover, given the noise and the low level of reliability in social media as opposed to patient reports in spontaneous reporting schemes, it is vital that we perform some form of preprocessing to guarantee the reliability of the identified ADRs.

In this paper we, therefore, consider the following problem: given a tweet *T* containing a drug *D* and an ADR *A*, whether *T* describes an instance where *A* is caused by *D*, as opposed to *A* and *D* co-occurring for a different reason (or randomly without any particular relation between *A* and *D*). Our experimental results show that the proposed method statistically significantly outperforms several baseline methods, demonstrating its ability to detect causality between drugs and ADRs in social media.

Related Work

The number of co-occurrences between a drug and an ADR can be used as a signal for detecting ADRs associated with drugs. Various measures have been proposed in the literature that evaluate the statistical significance of disproportionally large co-occurrences between a drug and an ADR. These include *Multiitem* Gamma Poisson Shrinker [24,26-28], Regression-Adjusted Gamma Poisson Shrinker [23], Bayesian Confidence Propagation Neural Network (BCPNN) [20-22], Proportional Reporting Rate [19,28], and Reporting Odds Ratio [19,28]. Each of these algorithms uses a different measure of disproportionality between the signal and its background. Information component is applied in BCPNN, whereas empirical Bayes geometric mean is implemented in all variants of the Gamma Poisson Shrinker algorithm. Each of the measures gives a specific score, which is based on the number of reports including the drug or the event of interest. These count-based methods are collectively referred to as *disproportionality measures*.

In contrast to these disproportionality measures that use only co-occurrence statistics for determining whether there is a positive association between a drug and an event, in this paper, we propose a method that uses the contextual information extracted from social media posts to learn a classifier that determines whether there is a causality relation between a drug and an ADR. Detecting causality between events from natural language texts has been studied in the context of discourse analysis [29,30] and textual entailment [31,32]. In discourse analysis, a discourse structure for a given text is created, showing the various discourse relationships such as causality, negation, and evidence. For example, in Rhetorical Structure Theory [33], a text is represented by a discourse tree where the nodes correspond to sentences or clauses referred to as elementary discourse units (EDUs), and the edges that link those textual nodes represent various discourse relations that exist between 2 EDUs. Supervised methods that require manually annotated discourse trees [34] as well as unsupervised methods that use discourse cues [35] and topic models [36] have been proposed for detecting discourse relations.

The problem of determining whether a particular semantic relation exists between 2 given entities in a text is a well-studied problem in the natural language processing (NLP) community. The context in which 2 entities co-occur provides useful clues for determining the semantic relation that exists between those entities. Various types of features have been extracted from co-occurring contexts for this purpose. For example, Cullotta and Sorensen [37] proposed tree kernels that use dependency trees. Dependency paths and the dependency relations over those paths are used as features in the kernel. Agichtein and Gravano [38] used a large set of automatically extracted surface-level lexical patterns for extracting entities and relations from large text collections.

To address the limitations of co-occurrence-based approaches, several prior studies have used contextual information [39]. Nikfarjam et al [40] annotated tweets for ADRs, beneficial effects, and indications and used those tweets to train a Conditional Random Field. They use contextual clues from tweets and word embeddings as features. Their problem setting is different from ours in the sense that we do not attempt to detect/extract ADRs or drug names from tweets but are only interested in determining whether the mentioned ADR is indeed relevant to the mentioned drug. A tweet can mention an ADR and a drug, but the ADR might not necessarily be related to the ADR. Huynh et al [41] proposed multiple deep learning models by concatenating convolutional neural network (CNN) and recurrent neural network architectures to build ADR classifiers. Specifically, given a sentence, they would like to create a binary classifier that predicts whether the sentence contains an ADR or otherwise. Their experimental results show CNNs to be the best for ADR detection. This observation is in agreement with broader text classification tasks in NLP where CNNs have reported the state-of-the-art performance [42]. However, one issue when using CNNs for ADR detection is the lack of labeled training instances, such as annotated tweets. This problem is further aggravated if we must learn embeddings of novel drugs or rare ADRs as part of the classifier training.

To overcome this problem, Lee et al [43] proposed a semisupervised CNN that can be pretrained using unlabeled data for learning phrase embeddings. Bidirectional Long Short-Term Memory (bi-LSTM) units were used [44] to tag ADRs and indicators in tweets. A small collection of 841 tweets was manually annotated by 2 annotators for this purpose. Pretrained word embeddings using skip-gram on 400 million tweets are used to initialize the bi-LSTM's word representations. This setting is different to what we study in this paper because we do not aim to tag ADRs and indicators in a tweet but to determine whether a tweet that mentions an ADR and a drug indicator describes an ADR event related to the drug mentioned in the tweet.

Methods

Overview

In this section, we presented our proposed method for detecting the causality between a drug and an event. First, in the section on "Problem Definition", we formally define the problem of causality detection between a drug and an event from social media posts. Next, we explain techniques for aggregating social media posts related to drugs and events. Next, we explain the method we use for extracting various lexical patterns that described the relationship between a drug and an event in social media posts. Finally, we present a machine learning approach that uses a manually annotated dataset containing social media posts as to whether they are describing a relationship between a drug and an adverse reaction for learning the reliability of the lexical patterns we extract. We have not assumed any specific properties or meta-data available in a particular type of social media platform such as *retweets*, *favorites* in Twitter, or *likes* or *comments* in Facebook. Although such platform-specific metadata can provide useful features for a machine learning algorithm, such metadata are not universally available across all social media platforms or cannot be retrieved due to privacy settings. The fact that the proposed method does not rely on such metadata was attractive because it made our proposed method applicable to a wide range of social media posts and does not limit it to a particular platform.

Problem Definition

Let us consider a social media post *T*, which explicitly mentions a drug *D* and an adverse reaction *R*. We model the problem of detecting causality between *D* and *R* in *T* as a binary classification problem where we would like to learn a binary classifier *h* (*T, D, R; w*) parametrized by a *d*-dimensional real-valued weight vector $w \in {}^{d}$ as shown in equation 1:

> (1) If *T* mentions that *D* causes *R*, then *h*(*T, D, R;w*)=1 and otherwise it is 0.

Here, we assume that the social media post *T* is already given to us and the drug and adverse reaction have already been detected in *T*. Detecting drug names can be done by matching against precompiled drug name lists (gazetteers) or using Named Entity Recognition [45]. A particular challenge when matching drug names in social media is that the drug names mentioned in social media might not necessarily match against the drug names listed in pharmacology databases [17]. The same drug is often sold under different labels by different manufacturers, and the label names continuously change, which makes it difficult to track a particular drug over time in social media. Similar challenges are encountered when matching ADRs in texts. Although the MedDRA [46] hierarchy assigns unique codes to preferred terms (PTs) that describe various ADRs such as "oropharyngeal swelling" or "systemic inflammatory response syndrome," such terms are used rarely by the majority of the social media users who might not necessarily be familiar with the MedDRA code names [47]. Although we acknowledge the challenges in detecting mentions of drug names and adverse reactions, we consider it to be beyond the scope of this paper, which focuses on a signal detection problem.

Social Media Aggregation

Although the problem definition described in Section 3.1 assumes that we are already provided with a set of social media posts, obtaining a large collection of social media posts relevant to drugs and events can be challenging for several reasons.

The vast majority of social media posts are not relevant to drugs or ADRs. One effective method for filtering out such irrelevant social media posts is to use the keyword-based filtering functionalities provided by the major social media application programming interfaces (APIs). As a specific example of such an API, we discuss the use of Twitter streaming API [48]. The Twitter streaming API allows registration of a set of keywords, and if there are any tweets that contain at least one of those keywords, then the corresponding tweet will be filtered and sent to the querying user. In our case, we used drug names and PTs (and their lexical variants) as keywords to filter the relevant tweets. Moreover, the streaming API also enabled us to limit the tweets to a particular geographical area or a language, which is useful if we want to monitor drugs that are specifically used in a particular country or a region.

Twitter's streaming API allowed us to aggregate tweets from 2 main types of data streams: *public streams* and *user streams*. Public streams are publicly available tweets by a specific group of users or on a topic. Hash tags in twitter are useful for streaming such public tweets on a particular topic. For example, by including the hash tag *#epilepsy*, we can retrieve tweets that are relevant to epilepsy. On the other hand, user streams allow us to obtain tweets from a single twitter user, containing roughly all of the data corresponding with that user's view (timeline) on Twitter. Despite the used aggressive filtering, streaming API returned a large number of tweets. Therefore, we stored the filtered tweets in a MongoDB [49] database in JavaScript Object Notation format for efficient retrieval.

Lexical Pattern Extraction

To represent the relationship between a drug and an ADR in a tweet, we extracted lexical patterns from the tweet. Let us illustrate the lexical pattern extraction process using the example tweet shown in Figure 2. We first identified the drug and event in the tweet and split the tweet into 3 parts. The part from the beginning of the tweet to the first-mentioned entity (either the drug or event) is named as the *prefix*, the part from the first-mentioned entity to the second-mentioned entity is named as the *midfix*, and the part from the second-mentioned entity to the end of the tweet is named as the *postfix*. Prior work on information extraction has shown that, in English, the midfix provides useful clues related to the relationship between 2 entities that co-occur in some context [50,51]. Indeed, from the example shown in Figure 2, we see that words such as *feeling* that appear in the midfix indicate that this twitter user is experiencing a side effect from the drug. However, it has also been shown that prefix and postfix terms also provide useful information when determining the relationship between 2 entities. For example, we see that the word *took* that appears in the prefix in the tweet (Figure 2), indicating that this twitter user has indeed taken this drug and not simply reporting an adverse reaction experienced by a different person. Such information is useful to estimate the reliability of the relationships mentioned in social media, which can often be noisy and unreliable. Therefore, in this work, we use all prefix, midfix, and postfix sections in tweets for extracting lexical patterns. We experimentally evaluate the significance of prefix, midfix, and postfix for ADR detection later in Section 4.

We extracted skip-grams from prefix, midfix, and postfix separately as lexical patterns for representing the relationship between a drug and an event. A skip-gram is an extension of *n*-gram. Unlike, *n*-grams that require us to consider all consecutive *n* words in a sequence, skip-grams allow us to generalize the *n*-gram patterns by skipping one or more words in a sequence. For example, trigram (*n*=3) lexical patterns extracted from the midfix shown in Figure 2 would be *while ago and*, *ago and now*, *and now feeling*, *now feeling very*.

On the other hand, skip-gram patterns also let us match any word (indicated by the wildcard "*") in an *n*-gram pattern. For example, the skip-gram pattern * *ago*, which is a generalization of the bigram pattern *while ago* will match various other time indicators such as *hours ago*, *days ago*, and *months ago*. Unlike, *n*-gram patterns that might not match exactly in numerous other tweets, skip-gram patterns flexibly match different tweets, thereby leading to a dense feature space. More importantly, skip-gram patterns subsume *n*-gram patterns. Therefore, all tweets that can be represented using *n*-gram patterns can be matched by the corresponding skip-gram patterns.

Considering the fragmented, ungrammatical, and misspelled texts frequently encountered in social media, skip-gram lexical patterns provide a robust and flexible feature representation. Moreover, extracting skip-grams is computationally efficient compared with, for example, part-of-speech tagging or dependency parsing social media, considering the volume of the texts we must process. Note that the drug name or the event are *not* part of the skip-gram lexical patterns. In other words, we replace the drug name and event, respectively, by place holder variables D and R. This is important because we would like to generate patterns that not only match the existing drugs and adverse reactions but can generalize to future drugs and their (currently unknown) adverse reactions. In our experiments, we use skip-gram lexical patterns for *n*=1, 2, and 3 and allowed a maximum of 1 wildcard in a pattern.

Learning Pattern Weights

We built a binary classifier that could predict whether an event *R* mentioned in a tweet *T* alongside a drug *D* was actually related to *D*. As explained later in Section 4.1, we used a manually annotated collection of tweets where each tweet contained a drug and an event, and a human annotator annotates whether the mentioned ADR is relevant to the drug (positively labeled instance) or otherwise (negatively labeled instance). We represent a tuple (T, D, R) using a feature vector $f\ (T, D, R) \in {}^{d}$, where each dimension corresponds to a particular skip-gram lexical pattern, which we extracted following the procedure described in Section 3.3. The value of the *i*-th dimension in the feature vector is set to 1 if the skip-gram lexical pattern $I_i l_i$, D, R) is represented by a boolean-valued feature vector over the set of skip-gram lexical patterns we extracted from all of the training instances. Using the above notation, let us denote this training dataset by $D_{train}=\{(f\ (T_n, D_n, R_n), y_n)\}^{N}{}_{n=1}$. Here, ($T_n$, D_n, R_n) indicates the *n*-th training instance out of *N* total instances in the dataset, and $y_n \in \{-1,+1\}$.

Unfortunately, not all skip-gram lexical patterns are equally important when determining whether there exists a relationship between a drug and an event. For example, in Figure 2, the pattern *while ago* can appear in various contexts, not necessarily in the context where an adverse reaction is described. Therefore, we assigned some form of a *confidence weight* to each skip-gram pattern before we used those patterns to make a decision about the relationship between a drug and an event. For this purpose, we assigned a weight *w* and *R* in *T* using the linear binary classifier given by equation 2:

$$(2)\ h(T, D, R; w)=\mathrm{sgn}(w^{T} f(T,D,R))$$

Here, $w \in {}^{d}$ is a *d*-dimensional real-valued weight vector where the *i*-th dimension represents the confidence weight w_i and the sign function sgn is defined in equation 3:

$$(3)\ \mathrm{sgn}(x)=1 \text{ if } x>0 \text{ and } -1 \text{ otherwise}$$

Figure 2. Extracting lexical patterns from a tweet that describes an adverse reaction (dizziness) caused by a drug (Atenolol). The tweet is split into 3 parts—prefix, midfix, and postfix, and various lexical patterns are extracted from each part. See text for the details of the pattern extraction method. Best viewed in color.

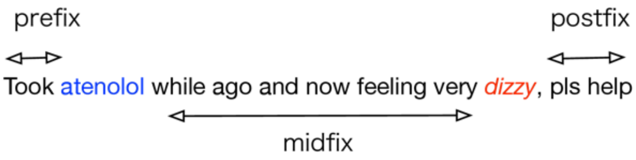


Figure 3. Support vector machine—optimization problem.

$$
\begin{aligned}
(4)\ \text{minimize}\quad & \frac{1}{2}||w||^2 + C\sum_{n=1}^{N}\xi_n \\
& y_n \boldsymbol{w}^\top f(T_n, D_n, R_n) \geq 1 \\
& \xi_n \geq 0
\end{aligned}
$$

Given the training dataset D_{train}, our goal was to learn w such that it can be used in equation 2 to predict whether the R mentioned in a T with D was indeed related to D. For this purpose, we used linear kernel support vector machines (SVMs) [52] with slack variables $\xi_n \geq 0$ noise in training instances. Given the scale of the annotation task, it is unavoidable that some of the instances will be incorrectly labeled by the human annotators, introducing some labeling noise to the training dataset. Second, slack variables can shift some of the training instances closer to the decision hyperplane, thereby artificially making the dataset to be linearly separable.

Although nonlinear kernels such as polynomial, radial basis function (RBF), or sigmoid can be used with SVMs, we limited our analysis to linear kernels for the following reason. Under the linear kernel, the weight associated with a particular feature can be seen as the influence imparted by that feature on the classification decision. This property is useful because we can identify the most discriminative lexical patterns that indicate a positive association between a drug and an event. We can use such lexical patterns, for example, to create extraction rules in the form of regular expressions to extract adverse reactions of drugs from social media. Because we are using a linear classifier in this work, it is important to handle the instances that violate the decision hyperplane using slack variables.

The joint learning of slack variables and weights can be formulated as the constrained convex optimization problem given by equation 4 in Figure 3.

Here, $C>0$, cost factor, is a hyperparameter that determines how much penalty we assigned to margin violations. The optimization problem given in equation 4 can be converted into a quadratic programming problem by introducing Lagrange multipliers. Efficient implementations that scale well to large datasets with millions of instances and features have been proposed [53].

Once we have obtained the weights w_i, equation 2 can be used to predict the relationship between D and R in T.

Results

We trained and evaluated the proposed method using a manually annotated dataset. The details of the dataset are presented in Section 4.1. Next, to evaluate the proposed method we compared it with several baseline methods. The baseline methods and their performances are described in Section 4.

To create a training and testing dataset for our task, we manually annotated a set of social media posts collected from the Twitter and Facebook between August 2015 and October 2015. Using the social media aggregation techniques described in Section 3.2, we filtered social media posts that contained a single mention of a drug and an event. The number of tweets that contain both a PT and a drug name was 94,890.

We then asked a group of annotators, who are familiar with ADRs of drugs, to annotate whether the event mentioned in the social media post is caused by the drug mentioned in the same post (a positively labeled instance) or otherwise (a negatively labeled instance).

The final annotated dataset contained 44,809 positively labeled instances and 50,081 negatively labeled instances. We perform 5-fold cross-validation on this dataset, selecting 80% of the positive and negative instances in each fold as training data, and the remainder as the testing data. In addition to the above-mentioned social media posts, we set aside 1000 positively and 1000 negatively labeled social media posts as developmental data, for tuning the hyperparameter C. In total, we extracted 168,663 skip-gram patterns from this dataset. We used classification accuracy defined by equation 5 as the evaluation measure:

> (5) Classification Accuracy=Total number of correctly predicted instances/Total number of instances in the dataset

Discussion

Baselines

We compared the proposed method with several baseline methods using the classification accuracy on the testing data as shown in Table 1. Next, we describe the different methods compared in Table 1.

Majority Baseline

Note that our training and test datasets were unbalanced in the sense that we have more negatively labeled instances than positively labeled instances. This situation is natural, given that most social media posts might not necessarily describe an adverse reaction of a drug even though it mentioned both the drug and an event. The training and test datasets we used in our evaluations closely simulate this situation. However, if a dataset is unbalanced, then by simply predicting the majority class (in our case this is the negative label) can still result in classification accuracies greater than 50%. The majority baseline shows the level of performance that was obtained by such a majority classifier.

Bag-of-Words Classifier

Our proposed method used skip-gram patterns for representing social media posts. An alternative approach would be to ignore the word order in the text and represent a text using the set of words contained in it. Specifically, we would represent each text by a binary-valued feature vector where the feature values for the unigrams that appear in the text are set to 1, and 0 otherwise. We then trained a binary SVM classifier with a linear kernel. By comparing against the bag-of-words (BOW) classifier, we can empirically evaluate the usefulness of the proposed skip-gram lexical patterns.

Prefix Only

This is a scaled-down version of the proposed method that used skip-gram patterns extracted only from the prefix. By evaluating against the prefix only baseline, we evaluated the importance of the information contained in the prefix. There are 50,021 prefix skip-gram patterns in total.

Midfix Only

This is a scaled-down version of the proposed method that uses skip-gram patterns extracted only from the midfix. By evaluating against the midfix only baseline, we evaluated the importance of the information contained in the midfix. There are 53,057 midfix skip-gram patterns in total.

Postfix Only

This is a scaled-down version of the proposed method that uses skip-gram patterns extracted only from the postfix. By evaluating against the postfix only baseline, we evaluated the importance of the information contained in the postfix. There are 65,585 postfix skip-gram patterns in total.

Prefix+Midfix

In this baseline method, we used both prefix and midfix for extracting skip-gram patterns. This baseline demonstrates the effectiveness of combining contextual information from both the prefix and the midfix.

Prefix+Postfix

In this baseline method, we used both prefix and postfix for extracting skip-gram patterns. This baseline demonstrates the effectiveness of combining contextual information from both the prefix and the postfix.

Midfix+Postfix

In this baseline method, we use both midfix and postfix for extracting skip-gram patterns. This baseline demonstrates the effectiveness of combining contextual information from the midfix and the postfix.

Convolutional Neural Network

We use the state-of-the-art short text classification method proposed by Kim [42] to train an ADR classifier. Each word in a tweet is represented using 128 dimensional word embeddings, where each dimension is randomly sampled from a uniform distribution in range (−1,1). The word embeddings are concatenated to represent a tweet. Next, a one-dimensional CNN with a stride size of 3 tokens and a max pooling layer is applied to create a fixed 20-dimensional tweet representation. We use Adaptive Subgradient Method [54] for optimization with initial learning rate set to 0.01 and the maximum number of iterations set to 1000. Finally, logistic sigmoid unit is used to produce a binary classifier.

Proposed Method

This is the method proposed in this paper. We use prefix, midfix, and postfix for extracting skip-gram patterns.

Using the development data, we found the cost parameter *C* for each setting. For the BOW classifier, the optimal *C* value was found to be 0.01, whereas for all the variants of the proposed method, it was 1.0.

The classification accuracies obtained for the 5-fold cross-validation task for the above-mentioned methods are shown in Table 1. From Table 1, we see that the majority baseline achieves an accuracy of 63.19%. Our task here is binary classification, and to compute confidence intervals for accuracies, we must compute binomial confidence intervals. There are several ways to compute this and one approach is the use of Clopper-Pearson confidence intervals [55]. By using confidence intervals, we can easily compare the statistical significance between methods, without having to conduct numerous pairwise comparisons between different methods. We compared all other methods against the accuracy reported by the majority baseline using Clopper-Pearson confidence intervals ($P<.001$) to test for statistical significance, which is (61.70,65.65). Statistically significant accuracies over the majority baseline are indicated by a superscripted letter a in Table 1.

From Table 1, we see that the best performance is obtained by the proposed method using the skip-gram patterns extracted from all prefix, midfix, and suffix contexts. A skip-gram pattern is an extension of n-gram patterns. Unlike n-gram patterns that must contain consecutive tokens, skip-gram patterns can skip one or more tokens when representing a subsequence. Among the different context types, we see that midfix performs best, whereas prefix and postfix perform relatively equally. This result is in agreement with prior work on information extraction for English, where midfix has been found to be useful. However, to the best of our knowledge, such an analysis has not yet been conducted for ADR extraction. Interestingly, we see that by adding the midfix to prefix and postfix we always perform better than if we had used only prefix or postfix. The proposed method uses all 3 contexts and obtains the best performance among the methods compared in Table 1. In particular, the performance reported by the proposed method is statistically significant over both the majority baseline and the BOW classifier. We see that the CNN-based ADR classifier is performing at the same level as the BOW classifier. Compared with the typical sentence classification datasets used to train such deep learning methods, our twitter dataset is significantly smaller, and this lack of data might have resulted in CNN-based ADR classifier to perform poorly in our experiments.

To gain further insights into the skip-gram patterns that are identified by the classifier to be useful for predicting whether there is a positive relationship between a drug and an event in a tweet, we plot the histogram of the feature weights in Figure 4. From Figure 4, we see that the majority of patterns have their weights close to zero, and an almost identical spread in positive and negative directions centered around zero. We counted 60,430 patterns to have weights exactly set to zero, meaning that approximately 35.83% (60,430/168,663) of patterns are found to be uninformative by the classifier. A randomly selected subset of zero-weighted patterns is shown in Table 1. Although there is a large number of patterns used as features, patterns that are not discriminative for the purpose of detecting ADRs are effectively pruned out by the SVM by assigning lower weights as shown in Table 2. Therefore, even if we have a comparatively larger feature space to the number of training instances, this does not necessarily result in overfitting.

We list the top-ranked positively weighted and negatively weighted skip-gram patterns in Table 3. From Table 3, we see that skip-gram patterns that describe a positive relationship between a drug and an ADR are correctly identified by the proposed method. For example, the *P+took+too* indicates that the user has actually took the drug. Moreover, we see many negations in the top-ranked negatively weighted patterns. Such clues could be used in several ways. First, we can use these clues as keywords for filtering social media posts that describe a potential positive relationship between drugs and ADRs. For example, we could run disproportionality-based signal detection methods using the disproportionality counts obtained from those filtered social media posts, thereby increasing the reliability of the detection. Second, these clues could be used to develop extraction patterns/templates that can be used for matching and extracting previously unknown ADRs for novel or existing drugs.

Table 1. Classification accuracy of different baselines and the proposed method.

Method	Classification accuracy
Majority baseline	63.19
Bag-of-words classifier	69.31[a]
Convolutional neural network	69.26[a]
Prefix only	66.41[a]
Midfix only	72.78[a]
Postfix only	68.08[a]
Prefix+midfix	74.72[a]
Prefix+postfix	71.07[a]
Midfix+postfix	77.10[a]
Proposed method	77.70[a]

[a] Statistically significant values.

Table 2. A randomly selected sample of features with zero weights.

Prefix patterns	Midfix patterns	Postfix patterns
P+trip+i	M+bad+idea	S+over
P+news+:	M+a+breakfast	S+12+hours
P+dat+lean	M+if+school	S+conquest
P+@rroddger	M+medica_authorities	S+please
P+fussiness+no	M+convicted+i	S+bad!

Table 3. Top-ranked positively (left 2 columns) and negatively (right 2 columns) weighted features (skip-gram patterns) by the support vector machine.

Feature	Weight	Feature	Weight
S[c]+als	1.2096	M+commercial	−1.2304
M[b]+induced	1.1314	P+hate+being	−1.0398
P[a]+oh+no	1.0683	P+I’m+definitely	−1.0000
M+[d]stinks	1.0000	P+clumsiness	−1.0000
S+.+wooh	1.0000	P+hospitalization	−1.000
M+never+work	1.0000	S+lol+fml	−0.9674
P+high+off	0.9006	S+wopps	−0.9035
P+took+too	0.8449	P+rt+xanaaxhadme	−0.8067
M+was+supposed	0.8378	P+don’t+think	−0.7721

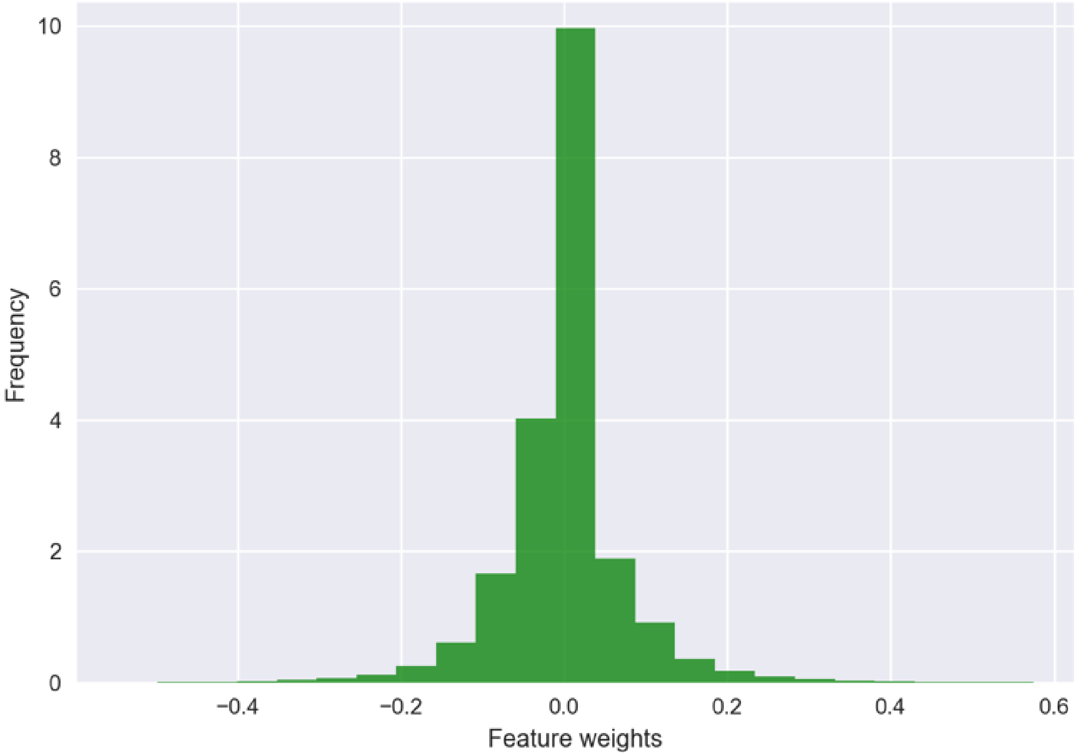
[a] P: prefix skip-gram patterns.

[b] M: midfix skip-gram patterns.

[c] S: postfix skip-gram patterns.

[d] For bigrams, we have used “+” to separate the constituent unigrams.

Figure 4. Histogram of the weights of the features learned by the support vector machine (SVM) classifier.



Conclusions

We proposed a novel signal detection problem where given a social media post *T* that contains a drug *D* and an event *R*, we would like to determine whether *R* is related to *D*, or otherwise. We have then proposed a method to solve this signal detection problem utilizing the lexical contextual information in *T*. Specifically, we extracted skip-gram patterns from the prefix, midfix, and suffix in *T*, and trained a binary SVM using a manually labeled training dataset. Our results show that the proposed method significantly outperformed the majority baseline and a BOW classifier. Moreover, we showed that the discriminative patterns were ranked at the top by the trained classifier. In the future, we plan to use the automatically extracted patterns to develop an ADR extraction method for previously unknown adverse reactions of drugs from social media.

Acknowledgments

All authors of this paper were supported by the WEB-RADR projected funded by the Innovative Medicines Initiative.

Conflicts of Interest

None declared.

References

1. Yang CC, Yang H, Jiang L, Zhang M. Social media mining for drug safety signal detection. 2012 Presented at: International Workshop on Smart Health and Wellbeing; October 29, 2012; Maui, Hawaii, USA p. 33-40.
2. Edwards IR, Aronson JK. Adverse drug reactions: definitions, diagnosis, and management. Lancet 2000 Oct 07;356(9237):1255-1259. [doi: 10.1016/S0140-6736(00)02799-9] [Medline: 11072960]
3. World Health Organization. International drug monitoring. The role of the hospital. World Health Organ Tech Rep Ser 1969;425:5-24. [Medline: 4981079]
4. Lee A. Adverse Drug Reactions. Glasgow, UK: Pharmaceutical Press; 2006.
5. Leaman R, Wojtulewicz L, Sullivan R, Skariah A, Yang J, Gonzalez G. Towards internet-age phramacovigilance: extracting adverse drug reactions from user posts to health-related social networks. 2010 Presented at: ACL Workshop on Biomedical Natural Language Processing; July 15, 2010; Uppsala, Sweden p. 117-125.
6. Giacomini KM, Krauss RM, Roden DM, Eichelbaum M, Hayden MR, Nakamura Y. When good drugs go bad. Nature 2007 Apr 26;446(7139):975-977. [doi: 10.1038/446975a] [Medline: 17460642]
7. Bates DW, Evans RS, Murff H, Stetson PD, Pizziferri L, Hripcsak G. Detecting adverse events using information technology. J Am Med Inform Assoc 2003;10(2):115-128 [FREE Full text] [Medline: 12595401]
8. van der Hooft CS, Sturkenboom MC, van Grootheest K, Kingma HJ, H Ch Stricker B. Adverse drug reaction-related hospitalisations. Drug Safety 2006;29(2):161-168. [doi: 10.2165/00002018-200629020-00006]
9. Rastegar-Mojarad M, Liu H, Nambisan P. Using social media data to identify potential candidates for drug repurposing: a feasibility study. JMIR Res Protoc 2016 Jun 16;5(2):e121 [FREE Full text] [doi: 10.2196/resprot.5621] [Medline: 27311964]
10. Yellow Card. URL: https://yellowcard.mhra.gov.uk/ [WebCite Cache ID 6yaAYCCVp]
11. Food and Drug Administration. Yellow Card System URL: https://www.fda.gov/Safety/MedWatch/ [WebCite Cache ID 6yaAaslWt]
12. Chaplin S. The yellow card scheme? Why are GPs under-reporting? Prescriber 2006;17(15):18-22 [FREE Full text]
13. Twitter. URL: https://twitter.com/ [accessed 2018-04-10] [WebCite Cache ID 6yaAg37Vl]
14. Facebook. URL: https://www.facebook.com/unsupportedbrowser [accessed 2018-04-10] [WebCite Cache ID 6yaAjjwQ4]
15. Instagram. URL: http://www.instagram.com/ [accessed 2018-04-10] [WebCite Cache ID 6yaAnyDIY]
16. Pinterest. URL: https://www.pinterest.com/ [accessed 2018-04-10] [WebCite Cache ID 6yaAlo8k8]
17. Sloane R, Osanlou O, Lewis D, Bollegala D, Maskell S, Pirmohamed M. Social media and pharmacovigilance: a review of the opportunities and challenges. Br J Clin Pharmacol 2015 Oct;80(4):910-920 [FREE Full text] [doi: 10.1111/bcp.12717] [Medline: 26147850]
18. WEB-RADR. URL: https://web-radr.eu/ [accessed 2018-04-10] [WebCite Cache ID 6yaApHoeV]
19. Suling M, Pigeot I. Signal detection and monitoring based on longitudinal healthcare data. Pharmaceutics 2012 Dec 13;4(4):607-640 [FREE Full text] [doi: 10.3390/pharmaceutics4040607] [Medline: 24300373]
20. Norén GN, Bate A, Orre R, Edwards IR. Extending the methods used to screen the WHO drug safety database towards analysis of complex associations and improved accuracy for rare events. Stat Med 2006 Nov 15;25(21):3740-3757. [doi: 10.1002/sim.2473] [Medline: 16381072]
21. Bate A, Lindquist M, Edwards IR, Orre R. A data mining approach for signal detection and analysis. Drug Safety 2002;25(6):393-397. [doi: 10.2165/00002018-200225060-00002]
22. Bate A, Lindquist M, Edwards IR, Olsson S, Orre R, Lansner A, et al. A Bayesian neural network method for adverse drug reaction signal generation. Eur J Clin Pharmacol 1998 Jun;54(4):315-321. [Medline: 9696956]
23. DuMouchel W, Harpaz R. Regression-adjusted gps algorithm (RGPS). Oracle Health Sci 2012 [FREE Full text]
24. Ahmed I, Haramguru F, Fourrier-Reglat A, Thiessard F, Kreft-Jias C, Miremont-Salame G, et al. Bayesian pharmacovigilance signal detection methods revisited in a multiple comparison setting. Stat Med 2009;28(13):1774-1792. [doi: 10.1002/sim.3586] [Medline: 19360795]
25. Fram DM, Almenoff JS, DuMouchel W. Empirical Bayesian data mining for discovering patterns in post-marketing drug safety. 2003 Presented at: Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; August 24-27, 2003; Washington, DC p. 359-368. [doi: 10.1145/956750.956792]
26. DuMouchel W, Pregibon D. Empirical bayes screening for multi-term associations. 2001 Presented at: International Conference on Knowledge Discovery and Data Mining; August 26-29, 2001; San Francisco, California. [doi: 10.1145/502512.502526]
27. Dumouchel W. Bayesian data mining in large frequency tables, with an application to the FDA spontaneous reporting system. Am Stat 1999;53(3):177-190. [doi: 10.2307/2686094]

28. Hauben M, Zhou X. Quantitative Methods in Pharmacovigilance. Drug Safety 2003;26(3):159-186. [doi: 10.2165/00002018-200326030-00003]
29. Do QX, Chan YS, Roth D. Minimally supervised event causality identification. 2011 Presented at: Empirical Methods in Natural Language Processing; July 27-31, 2011; Edinburgh, UK.
30. Radinsky K, Davidovich S, Markovitch S. Learning causality for news events prediction. 2012 Presented at: 21st International Conference on World Wide Web; April 16-20, 2012; Lyon, France p. 909-918.
31. Androutsopoulos I, Malakasiotis P. A survey of paraphrasing and textual entailment methods. J Artif Intell Res 2010;38:135-187. [doi: 10.1613/jair.2985]
32. Angel Ríos Gaona M, Gelbukh A, Bandyopadhay S. Recognizing textual entailment with statistical methods. 2010 Presented at: 2nd Mexican Conference on Pattern Recognition: Advances in Pattern Recognition; September 27-29, 2010; Puebla, Mexico p. 372-381.
33. Mann WC, Thompson SA. Rhetorical structure theory: toward a functional theory of text organization. Text Talk 1988;8(3):243-281. [doi: 10.1515/text.1.1988.8.3.243]
34. duVerle DA, Prendinger H. A novel discourse parser based on support vector machine classification. 2009 Presented at: Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP; August 2-7, 2009; Suntec, Singapore p. 665-673.
35. Marcu D, Echihabi A. An unsupervised approach to recognizing discourse relations. 2002 Presented at: Annual Conference of the Association for Computational Linguistics; July 7-12, 2002; Philadelphia, Pennsylvania p. 368-375.
36. Ó Séaghdha D, Teufel S. Unsupervised learning of rhetorical structure with un-topic models. 2014 Presented at: 25th International Conference on Computational Linguistics; August 23-29, 2014; Dublin, Ireland p. 2-13 URL: http://www.aclweb.org/anthology/C14-1002
37. Culotta A, Sorensen J. Dependency Tree Kernels for Relation Extraction. 2004 Presented at: 42nd Annual Conference of the Association for Computational Linguistics; July 21-26, 2004; Barcelona, Spain p. 423-429 URL: http://aclweb.org/anthology/P/P04/P04-1054.pdf?CFID=28410093&CFTOKEN=4e679354be37e14d-B170E708-9B8A-B137-2D13DA2B3C10E01A [doi: 10.3115/1218955.1219009]
38. Agichtein E, Gravano L. Snowball: extracting relations from large plain-text collections. 2000 Presented at: Fifth ACM International Conference on Digital Libraries; June 2-7, 2000; San Antonio, Texas, USA p. 85-94. [doi: 10.1145/336597336644]
39. Lardon J, Abdellaoui R, Bellet F, Asfari H, Souvignet J, Texier N, et al. Adverse drug reaction identification and extraction in social media: a scoping review. J Med Internet Res 2015 Jul 10;17(7):e171 [FREE Full text] [doi: 10.2196/jmir.4304] [Medline: 26163365]
40. Nikfarjam A, Sarker A, O'Connor K, Ginn R, Gonzalez G. Pharmacovigilance from social media: mining adverse drug reaction mentions using sequence labeling with word embedding cluster features. J Am Med Inform Assoc 2015 May;22(3):671-681 [FREE Full text] [doi: 10.1093/jamia/ocu041] [Medline: 25755127]
41. Huynh T, He Y, Willis A, Rueger S. Adverse drug reaction classification with deep neural networks. 2016 Presented at: International Conference on Computational Linguistics; December 11-17, 2016; Osaka, Japan URL: http://www.aclweb.org/anthology/C16-1084
42. Kim Y. Convolutional neural networks for sentence classification. 2014 Presented at: Conference on Empirical Methods in Natural Language Processing; October 25-29, 2014; Doha, Qatar p. 1746-1751 URL: http://www.aclweb.org/anthology/D14-1181
43. Lee K, Qadir A, Hasan S, Datla V, Prakash A, Liu J, et al. Adverse Drug Event Detection in Tweets with Semi-Supervised Convolutional Neural Networks. 2017 Presented at: International World Wide Web Conference; April 3-7, 2017; Perth, Australia p. 705-714. [doi: 10.1145/3038912.3052671]
44. Cocos A, Fiks AG, Masino AJ. Deep learning for pharmacovigilance: recurrent neural network architectures for labeling adverse drug reactions in Twitter posts. J Am Med Inform Assoc 2017 Jul 01;24(4):813-821. [doi: 10.1093/jamia/ocw180] [Medline: 28339747]
45. Segura-Bedmar I, Martínez P, Segura-Bedmar M. Drug name recognition and classification in biomedical texts. A case study outlining approaches underpinning automated systems. Drug Discov Today 2008 Sep;13(17-18):816-823. [doi: 10.1016/j.drudis.2008.06.001] [Medline: 18602492]
46. MedDRA. URL: https://www.meddra.org/ [accessed 2018-04-10] [WebCite Cache ID 6yaAu2y44]
47. Limsopatham N, Collier N. Adapting phrase-based machine translation to normalise medical terms in social media messages. 2015 Presented at: Conference on Empirical Methods in Natural Language Processing; September 17-21, 2015; Lisbon, Portugal p. 1675-1680 URL: https://aclanthology.info/pdf/D/D15/D15-1194.pdf
48. Twitter. Twitter Streaming API URL: https://developer.twitter.com/en/docs [WebCite Cache ID 6yaAwaFiY]
49. MongoDB. URL: https://www.mongodb.com/ [accessed 2018-04-10] [WebCite Cache ID 6yaAzTZdT]
50. Bollegala D, Matsuo Y, Ishizuka M. A relational model of semantic similarity between words using automatically extracted lexical pattern clusters from the web. 2009 Presented at: Empirical Methods in Natural Language Processing; August 6-7, 2009; Singapore p. 803-812.

51. Bollegala D, Matsuo Y, Ishizuka M. Measuring the similarity between implicit semantic relations from the web. 2009 Presented at: International World Wide Web Conference (WWW 2009); April 20-24, 2009; Madrid, Spain p. 651-660. [doi: 10.1145/1526709.1526797]
52. Vapnik VN. Statistical Learning Theory. Chichester, UK: Wiley; 1998.
53. Fan RE, Chang KW, Hsieh CJ, Wang XR, Lin CJ. LIBLINEAR: A library for large linear classification. J Mach Learn Res 2008:1871-1874.
54. Duchi J, Hazan E, Singer Y. Adaptive subgradient methods for online learning and stochastic optimization. J Mach Learn Res 2011;12:2121-2159 [FREE Full text]
55. Clopper CJ, Pearson ES. The use of confidence or fiducial limits illustrated in the case of the binomial. Biometrika 1934 Dec;26(4):404. [doi: 10.2307/2331986]

Abbreviations

ADRs: adverse drug reactions
API: application programming interface
BCPNN: Bayesian Confidence Propagation Neural Network
Bi-LSTM: Bidirectional Long Short-Term Memory
BOW: bag of words
CNN: convolutional neural network
EDUs: elementary discourse units
NLP: natural language processing
PTs: preferred terms
SVMs: support vector machines

Edited by G Eysenbach; submitted 15.06.17; peer-reviewed by N Okazaki, N Limsopatham; comments to author 22.09.17; revised version received 25.10.17; accepted 14.03.18; published 09.05.18.

Please cite as:
Bollegala D, Maskell S, Sloane R, Hajne J, Pirmohamed M
Causality Patterns for Detecting Adverse Drug Reactions From Social Media: Text Mining Approach
JMIR Public Health Surveill 2018;4(2):e51
URL: http://publichealth.jmir.org/2018/2/e51/
doi:10.2196/publichealth.8214
PMID:29743155

Original Paper

Estimating the Risk of Influenza-Like Illness Transmission Through Social Contacts: Web-Based Participatory Cohort Study

Ta-Chien Chan[1], PhD; Tsuey-Hwa Hu[2], MSc; Jing-Shiang Hwang[2], PhD

[1]Research Center for Humanities and Social Sciences, Academia Sinica, Taipei City, Taiwan
[2]Institute of Statistical Science, Academia Sinica, Taipei City, Taiwan

Corresponding Author:
Jing-Shiang Hwang, PhD
Institute of Statistical Science
Academia Sinica
128 Academia Road, Section 2
Taipei City, 115
Taiwan
Phone: 886 2 2783 5611
Fax: 886 2 2783 1523
Email: hwang@sinica.edu.tw

Abstract

Background: Epidemiological studies on influenza have focused mostly on enhancing vaccination coverage or promoting personal hygiene behavior. Few studies have investigated potential effects of personal health behaviors and social contacts on the risk of getting influenza-like illness (ILI).

Objective: Taking advantage of an online participatory cohort, this study aimed to estimate the increased risk of getting ILI after contact with infected persons and examine how personal health behaviors, weather, and air pollution affect the probability of getting ILI.

Methods: A Web-based platform was designed for participants to record daily health behaviors and social contacts during the influenza season of October 1, 2015 to March 31, 2016, in Taiwan. Data on sleep, diet, physical activity, self-reported ILI, and contact with infected persons were retrieved from the diaries. Measurements of weather and air pollutants were used for calculating environmental exposure levels for the participants. We fitted a mixed-effects logistic regression model to the daily measurements of the diary keepers to estimate the effects of these variables on the risk of getting ILI.

Results: During the influenza season, 160 participants provided 14,317 health diaries and recorded 124,222 face-to-face contacts. The model estimated odds ratio of getting ILI was 1.87 (95% CI 1.40-2.50) when a person had contact with others having ILI in the previous 3 days. Longer duration of physical exercise and eating more fruits, beans, and dairy products were associated with lower risk of getting ILI. However, staying up late was linked to an elevated risk of getting ILI. Higher variation of ambient temperature and worse air quality were associated with increased risk of developing ILI.

Conclusions: Developing a healthier lifestyle, avoiding contact with persons having ILI symptoms, and staying alert with respect to temperature changes and air quality can reduce the risk of getting ILI.

(JMIR Public Health Surveill 2018;4(2):e40) doi:10.2196/publichealth.8874

KEYWORDS

flu transmission; social networks; contact diary; diet; exercise; sleep quality

Introduction

Background

Seasonal influenza epidemics cause a high disease burden, including direct costs of health services and households and indirect costs due to productivity losses worldwide every year [1-3]. Preventing influenza transmission includes pharmaceutical approaches [4,5] and nonpharmaceutical approaches [6]. Human face-to-face contact is the major transmission route for influenza, and infection probability will be elevated with close contact [7,8]. The household is one of the places where people have close contact and is also the bridge of influenza transmission between schools and the community [9]. If one member of a household gets influenza, the risk of infection in a household contact can be up to 38% [10]. Although social contact enhances the risk of influenza infection, there are still some ways to break down this transmission route, including personal prevention and avoiding social contact. In personal prevention, enhancing personal immunity through physical exercise and keeping good hygiene behaviors such as hand washing and wearing a face mask are effective approaches to reduce the risk of contracting an influenza infection [11]. Avoiding face-to-face contact with infected people should also reduce the chance of getting an infection. However, it has been a challenge to prospectively quantify how social contact affects the chance of influenza transmission, including the number of instances, intensity, and duration of contacts [8]. In one social network study, the researchers found that people at the center of a network got the flu earlier, and that information could be used for early detection of influenza epidemics [12]. This implies that people with more social ties might have higher chances of getting infection or spreading the virus to others. In addition to those risk factors from human beings, external weather conditions not only facilitate influenza transmission but also affect people's immunity [13,14]. Recently, one study reported that increased exposure to ambient fine particulate matter ($PM_{2.5}$) also contributed to 10.7% of incident influenza cases in China [15]. From observation of human cell lines, exposure to ozone also is associated with increased influenza susceptibility [16].

Objectives

In this study, we retrieved data from an online diary platform called ClickDiary [17], which provided participants a simple way to prospectively record daily health behaviors, influenza symptoms in themselves, and their contact persons during an influenza season in Taiwan. There have previously been some studies using self-reporting of influenza-like illness (ILI) by volunteers to monitor influenza activity in the communities, such as the *Flutracking* platform in Australia [18] and *Flu Near You* in the United States and Canada [19], both of which are simple internet-based surveys to record the weekly ILI symptoms of the participants. However, the main purpose of those studies was to detect earlier aberration signals to detect flu epidemic. In contrast, with the comprehensive diary data, the purpose of this study was to estimate the increased risk of getting ILI symptoms after contact with infected persons and examine how personal health behaviors, weather, and air pollution affect the chance of getting ILI.

Methods

Ethics

This study was approved by the Institutional Review Board on Humanities and Social Science Research, Academia Sinica (AS-IRB-HS 02-13022). The diary data for analysis were stripped of personal identification information, which was replaced with a serial number to protect participants' privacy.

ClickDiary Program

We used a Web-based platform named "ClickDiary" [20] to collect data on participants' daily health behaviors and social contact [17]. The participants, recruited from various channels, included university students, school teachers and administrative employees, volunteers at health promotion centers, hospital patients, and community college students, as well as other adults in the general population. This ClickDiary program started in May 2014 and ended in December 2016. Although 1432 participants contributed their diaries, active diary keepers numbered about 200 in any given week. The participants were required to click questionnaire options in the online diary at least twice a week. In the health diary, the participants recorded sleep duration and quality, mood, food intake, physical exercise, self-reported ILI and symptoms, and other information. The diet questionnaire is adopted and modified from an official diet questionnaire created by the Health Promotion Administration, Ministry of Health and Welfare, Taiwan. The exercise questionnaire is adopted from one health examination questionnaire [21]. The detailed diet and exercise questionnaires were listed in another published paper [17]. The dietary intake consisted of 14 major food items, and the participants were to report the number of portions (0, 0.5, 1, 2, 3, or 3+) consumed during the past 24 hours. Food items included vegetables, fruits, whole grains and roots, rice and flour, pork/beef/mutton, chicken/duck/goose, seafood, eggs, beans and pulses, dairy products, fried foods, processed foods, desserts, and sugary drinks. The participant reported whether he or she had ILI by selecting one of the 3 options, including none, probably, and definitely having ILI symptoms. The participants reporting ILI were also asked to select at least one of the specific symptoms, including fever (>38°C), cough, sneezing, chills, stuffy nose, sore throat, fatigue, diarrhea, and/or chest pain. In this study, we defined a participant as having ILI symptoms only when she or he selected "definitely having ILI symptoms."

The contact diary was designed for collecting information on participants' daily social contact. The definition of contact used was stricter than in other contact diary studies [22]. An instance of contact was defined as in-person exchange of at least three sentences [23].The participants recorded major contact attributes, including time and place of contact, who initiated it, the means of contact, duration, content, participant's instrumental gain and mood, and the contact person's location, mood, and ILI symptoms, if any. Each record in the contact diary represented a one-on-one contact. If the participant met many people at the same time, they recorded only the persons with whom they had exchanges of at least three sentences. A lottery-based reward system was designed to encourage the participants to fulfill the requirements. To ensure the quality of the data, we checked the data pattern of each participant every week. If a participant had not kept his or her diaries properly, the participant was put on an alert list, and any data entered by the participant would be excluded from the database for analysis.

The participants provided demographic information at sign-up, including age, gender, place of residence, marital status, and type of current job. The program also collected participants' Big Five personality traits (openness, conscientiousness, extraversion, agreeableness, and neuroticism) [24], height and weight for calculating body mass index (kg/m^2), perceived health status and happiness, the number (and characteristics) of people with whom they were in contact during the day, and a baseline health survey, which borrows items from the Taiwan Social Change Survey [25].

Diary Data and Notation

We created a cohort of 160 participants who each recorded the two diaries at least 10 days a month for more than 2 months during the influenza season from October 1, 2015, to March 31, 2016. Some participants recorded their health diary most of the times, but their contact diary only occasionally, because it was more time-consuming. Although the participants were required to input in their diaries twice a week, we found that more than half of the participants in the cohort had recorded entries at least 3 times a week, and the average was 4 times a week. There are two major reasons why we did not include all data across the entire period. The first one is that our data showed that the overall self-reported ILI rate was low in the nonflu season. To focus on identifying the risk factors for ILI transmission, we restricted the studied period to the flu season. The other reason is that, for the first 2014-2015 flu season, we encountered a logistical problem for issuing the vouchers from mid-December 2014 to January 2015. This caused the participation rate to fall during that period, and the number of diaries was too small to conduct the analysis.

We introduce a notation to describe the variables for use in the statistical analysis in the following. When the *i*-th participant recorded the health diary for the *j*-th time on the day t_{ij} during the study period, we denoted $Y_{ij}=1$ if the participant reported ILI symptoms and 0 otherwise. We then retrieved data logs of the participant's health diary and contact diary for the 3 days before t_{ij}. The binary variables $Z_{ij}=1$ and $C_{ij}=1$represent the participant self-reporting ILI and having contact with infected persons, respectively, during the past 3 days, and 0 otherwise. We count events in the previous 3 days based on the incubation period of influenza [10]. For each food item, the number of portions 3+ was coded as 4. The average number of portions of the *k*-th food category in the past 3 days is denoted by F_{kij}. The participants' selected daily exercise time from five categories ranging from none to 1-30, 31-60, 61-120, and 120+ minutes. We coded these 5 categories as 0, 15.5, 45.5, 90.5, and 120 min [26]. The average duration of exercise in the past 3 days is denoted by E_{ij}. We denoted the average sleep time of the participants as S_{ij}, and staying up after 1 AM at least once in the past 3 days as L_{ij}.

Weather and Air Pollution Data

To understand how environmental factors affect influenza transmission after controlling for health behaviors and social contacts, we included weather and air pollution data from ambient air quality monitoring stations maintained by Taiwan's Environmental Protection Administration, excluding traffic, industrial, and background stations [27]. In this study, we selected 35 air monitoring stations closest to the townships the participants lived in. The measurements of weather and air pollutants at the selected stations were used for computing environmental exposure for the participants. We computed average values of daily mean temperature (°C), daily maximum 8-hour moving average concentration of O_3 (ppm), daily mean concentration of $PM_{2.5}$ (μg/m^3) and other pollutants, and SD of daily mean temperature (°C) in the past 3 and 7 days.

Statistical Analysis

In addition to the influence of contact with infected persons and self-reported ILI status in the past 3 days, as well as age and gender, we first applied logistic regression models to identify health behaviors associated with the probability of self-reported ILI according to the following equation:

Specifically, we fit the logistic regression model for i=1,2,…,160 and j=1,2,…, D_i, where $G_i=1$ indicates the subject is male, A_i is age in years, and D_i is the number of diary entries the *i*-th participant provided. The coefficientδ_{01} is one of the relevant parameters representing the effect of contact with infected persons and no self-reported ILI in the past 3 days on risk of having ILI symptoms on the current day.

The explanatory variables were selected into the model in two stages. At the first stage, we considered personal risk factors, including age, gender, contact with persons with ILI or not, average portions of different kinds of foods in the past 3 days or in the past 7 days, staying up late, average number of hours of sleep, sleeping quality, average mood scores, and amount of exercise in the past 3 days or in the past 7 days. The likelihood ratio test was used to select important variables. We then retain the identified influential health behavior variables in the logistic regression models. In the second stage, we continue by identifying influential weather including temperature and humidity and air pollution variables, including $PM_{2.5}$, SO_2, O_3, and CO using a stepwise approach. Each variable was computed in two temporal windows. One was for the past 3 days and the other was for the past 7 days. When all potential covariates were fixed in the logistic regression model equation, we further added a random component to the intercept for modeling subject-to-subject variation. Because the influence on the influenza risk of a reported portion of food intake may have been different among the participants, we also added random components to the coefficients of the chosen food items in the regression equation. Finally, we assume that the repeated records of each participant are correlated in the model. All the added random components were assumed to be normally distributed with mean 0 and constant variance. Because each participant provided self-reported ILI status repeatedly during the study period, we further assumed that a pair of responses of a subject had the correlation according to the following equation;

where Y_{ij} is the response of subject *i* on day t_{ij} and δ is a positive parameter to be estimated. We used the R software (R Foundation for Statistical Computing, Vienna, Austria; version 3.3.2) [28] and R package MASS [29] using the glmmPQL function to estimate the parameters in the final mixed-effects logistic regression model.

Results

Exploratory Data Analysis

The mean, minimum, and maximum number of days on the participants recorded entries in the health diary were 89, 29, and 183 days, respectively, and for recording the face-to-face contact diary, the same statistics were 91, 6, and 183 during the study period of 183 days, whereas for days with an entry in either diary, the same statistics were 99, 29, and 183 (Table 1). Consequently, we were able to retrieve 14,317 person-days for the health diary and 124,222 face-to-face contacts from the cohort for analysis. The self-reported data from the participants gave an incidence rate estimate of 3.19% (456 ILI persons/14,317 person-days) for the 2015-2016 influenza season. The diary keepers reported 1045 face-to-face contacts with persons definitely having ILI among 124,222 face-to-face contacts, or a chance of 0.84% (1045/124,222), in this influenza season.

A total of 160 participants were included in this study (Table 2). Most of the participants were female (122/160, 76.3%) and tended to be young adults, aged between 20 and 40 years (99/160, 61.9%). When we considered face-to-face contacts during the previous 3 days, there were 1806 contacts on a given day without any additional contact records from the contact diary for the previous 3 days. Thus, we removed those records from the original 14,317 person-days, and the final total number of person-days was 12,251. Among those contacts (Table 3), for only 131 contacts (131/12,251, 1.07%), both participants and their contacted alters had ILI symptoms during the previous 3 days. For most contacts (10,974/12,251, 89.58%), both participants and their contacted alters were free of ILI. The time series of the daily incidence rate of self-report ILI and the weekly outpatient ILI admission rate plotted in Figure 1 show that the self-reported data of the cohort matched the actual nationwide outpatient surveillance data, especially around the peak, obtained from Taiwan's Centers for Disease Control. The Pearson correlation was .74 between the weekly incidence rate of self-reported ILI and the weekly outpatient ILI admission rate.

After the first stage of selecting influential variables from using the logistic regression models, there were 12 variables retained in the final model. The descriptive statistics of the identified variables are listed in Table 4. Most of the variables selected here were calculated using the measurements in the prior 3 days, except for the temperature deviation, which used the past 7 days. A total of 12,251 records were included in the final model. The average daily servings of vegetables, fruits, beans and pulses, and dairy products ranged from 0 to 4. The category of cereals comprised two subcategories of food including grains and root vegetables, and rice and noodles, portions of which ranged from 0 to 8 servings. The category of meats and egg contained four subcategories of food, including red meats, white meats, fish, and eggs, which ranged from 0 to 14.5 servings. The proportion of the sleep records in health diaries indicating staying up late (till 1 AM) was 24.96% (3,058/12,251). The average sleeping time was 7.4 hours, with a range from 0.5 to 14.3 hours. The average exercise time was 21 min, ranging from 0 to 120 min. The average temperature deviation in the past 7 days was 2.16°C, ranging from 0.12°C to 5.94°C. The average of maximum $PM_{2.5}$ in the past 3 days was 30 $\mu g/m^3$, ranging from 2 to 88.6 $\mu g/m^3$. The average of maximum 8-hour moving average of O_3 was 35 ppb, which ranged from 5.7 to 76.2 ppb.

Table 1. Descriptive statistics of number of days when participants recorded health diary and face-to-face contact diary, from October 1, 2015 to March 31, 2016.

Type of diary	Number of days					
	Minimum	25%	50%	75%	Maximum	Mean
Health diary	29	63	76	113	183	89
Contact diary	6	64	71	116	183	91
Either diary	29	66	91	128	183	99

Table 2. Demographic summary of 160 participants.

Age group	Gender		All
	Male, n (%)	Female, n (%)	n (%)
20-30	14 (36.8)	45 (36.9)	59 (36.9)
31-40	10 (26.3)	30 (24.6)	40 (25.0)
41-50	6 (15.8)	20 (16.4)	26 (16.3)
51-60	2 (5.3)	19 (15.6)	21 (13.1)
61-70	6 (15.8)	8 (6.6)	14 (8.8)
All	38 (100.0)	122 (100.0)	160 (100.0)

Table 3. Number and percentage of influenza-like illness (ILI) between participants and their contact persons.

Participants had ILI in past 3 days	Contact persons had ILI in past 3 days	
	No, n (%)	Yes, n (%)
No	10,974 (89.58)	824 (6.72)
Yes	322 (2.63)	131 (1.07)

Figure 1. Temporal trends of daily and weekly incidence rate of self-report influenza-like illness (ILI) and weekly outpatient ILI admission rate.

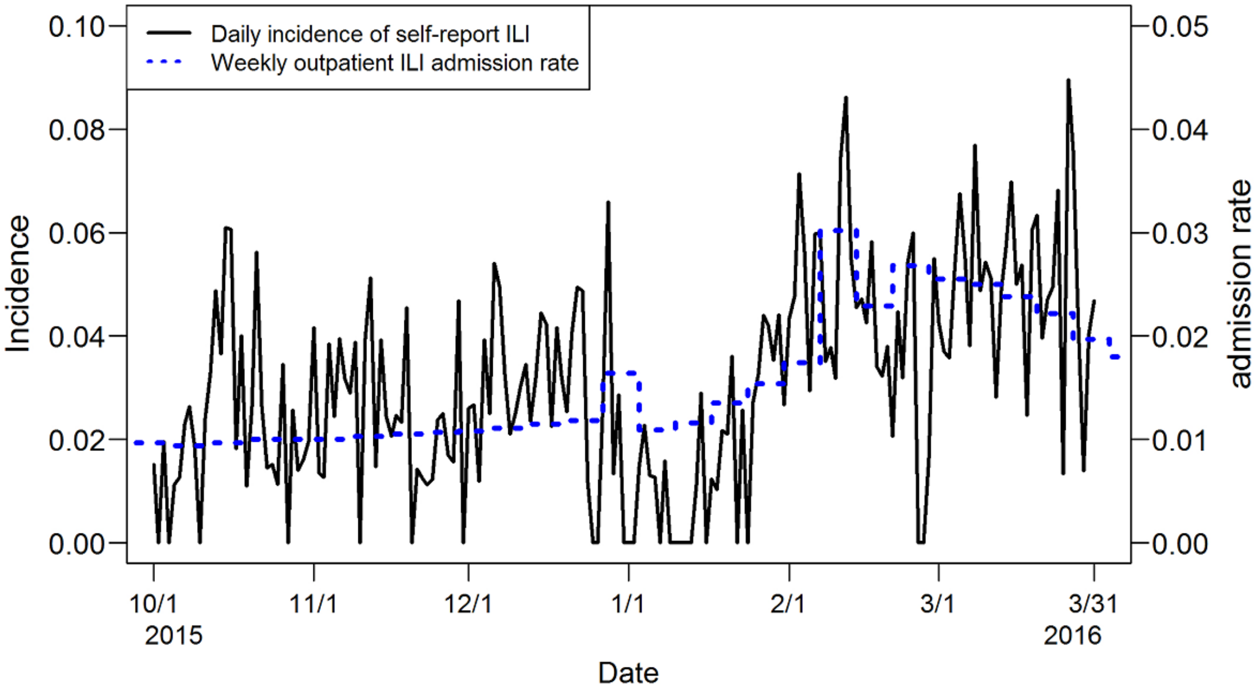


Table 4. Descriptive statistics of selected variables for fitting the models. Q1 and Q3 represent the first and third quartiles.

Variables	Descriptive statistics					
	Minimum	Q1	Median	Mean	Q3	Maximum
Diet (portion)						
Vegetables	0	1	2	1.75	2	4
Fruits	0	0.5	1	1.28	2	4
Cereals	0	1.75	2.5	2.55	3	8
Beans and pulses	0	0	0.5	0.61	1	4
Meats and eggs	0	1	2	2.34	3.17	14.5
Dairy products	0	0	0.25	0.4	0.67	4
Sleep time (hours)	0.5	6.5	7.33	7.38	8.25	14.25
Exercise (min)	0	0	15.5	21.44	30.5	120
Temperature deviation (°C)[a]	0.12	1.38	1.92	2.16	2.76	5.94
Maximum $PM_{2.5}$ ($\mu g/m^3$)	2.04	19.25	27.29	30.01	38.02	88.61
Maximum 8-hour moving average of O_3 (ppb)	5.69	28.17	33.76	34.98	40.78	76.22

[a] Only temperature deviation represented values in the past 7 days; other variables represented values in the past 3 days.

Table 5. Estimates of the odd ratios in the mixed-effects logistic regression models. IQR: interquartile range; OR: odds ratio; ILI: influenza-like illness.

Variables	IQR	OR (95% CI)
Binary variables		
Free of ILI and contact with infected persons[a]		1.87 (1.40-2.50)
Self-reporting ILI and no contact with infected persons[a]		55.79 (45.26-68.77)
Self-reporting ILI and contact with infected persons[a]		59.97 (44.32-81.14)
Age >60[b]		0.06 (0.0005-8.23)
Male[c]		0.30 (0.05-1.76)
Late bedtime[d]		1.43 (1.11-1.84)
Continuous variables		
Vegetables	1.0	0.92 (0.64-1.33)
Fruits	1.5	0.37 (0.19-0.75)
Cereals	1.25	0.99 (0.70-1.40)
Beans and pulses	1.0	0.42 (0.20-0.87)
Meats and eggs	2.17	1.09 (0.67-1.77)
Dairy products	0.67	0.31 (0.14-0.69)
Sleep duration (h)	1.67	0.97 (0.84-1.12)
Exercise time	30.5	0.73 (0.63-0.84)
Temperature deviation	1.37	1.25 (1.13-1.39)
log ($PM_{2.5}$)[e]	0.68	1.13 (0.99-1.30)
O_3	12.66	1.33 (1.20-1.49)
For two continuous variables		
log ($PM_{2.5}$) and O_3	0.68 and 12.66	1.51 (1.29-1.76)

[a] Reference group: Free of ILI and no contact with infected persons.

[b] Reference group: Age≤60.

[c] Reference group: Female.

[d] Reference group: Did not have late bedtime.

[e] $PM_{2.5}$: fine particulate matter.

Model Results

The fixed-effects logistic regression models identified several influential variables associated with the probability of participants' reporting ILI symptoms. The variables include reporting having had any of the following during the past 3 days: contacts with infected persons; being free of ILI; staying up late; average exercise time; and average consumption of fruits, beans and pulses, and dairy products. The most influential environmental variables identified were SD of daily mean temperature in the past 7 days, mean daily maximum 8-hour moving average ozone, and mean daily $PM_{2.5}$ concentrations in the past 3 days. The estimated coefficients of the model shown in Table 5 indicate that the odds ratio (OR) of reporting ILI symptoms was 1.87 with a 95% CI (1.40-2.50) when the participants had contact with persons with ILI symptoms in the past 3 days (hereafter, "recent"). The model estimates indicated that recent consumption of more fruits, beans and pulses, and dairy products significantly decreased the risk of getting ILI. The estimated OR with 95% CI was 0.37 (95% CI 0.19-0.75) for comparing the interquartile range (IQR) of a 1.5-portion increase of daily fruit intake. For recent daily consumption of beans and pulses, the estimated OR was 0.42 (95% CI 0.20-0.87) for comparing an IQR of 1 portion. Comparing an IQR of a two-thirds portion of dairy products, the estimated OR was 0.31 (95% CI 0.14-0.69). Spending more time exercising recently also reduced the risk; the estimated OR was 0.73 (95% CI 0.63-0.84) for comparing an IQR of 30 min. However, staying up later than 1 AM during any night in the past 3 days was associated with increased risk of reporting ILI symptoms, with an estimated OR of 1.43 (95% CI 1.11-1.84).

Variation of daily mean temperature in the past 7 days also increased the risk of infection. The estimated OR was 1.25 (95% CI 1.13-1.39) for comparing an IQR of 1.37°C in SD of the daily temperatures in the past 7 days. Participants exposed to higher ozone concentrations in the past 3 days had a higher chance of reporting ILI symptoms. The estimated OR was 1.33 (95% CI 1.20-1.49) for an increased IQR of 12.7 ppm in the daily maximum 8-hour moving average concentration of ozone.

The concentration of $PM_{2.5}$ on a log scale was found to be only marginally associated with response from the fitted model. The estimated OR was 1.13 (95% CI 0.99-1.30) for comparing an IQR of 18.8 μg/m^3 in daily average concentration of $PM_{2.5}$. Because the two variables of ozone and $PM_{2.5}$ had a correlation of .42, we compared the third quartile concentrations of both ozone and $PM_{2.5}$ against their first quartile levels. The estimated OR increased to 1.5 (95% CI 1.29-1.76). When the participants reported ILI symptoms in the past 3 days, we expected large influence on reporting ILI on the current day even if they had no contact with infected persons. The model estimated OR was 55.8 (95% CI 45.3-68.8).

We removed the insignificant variables of vegetables, cereals, meats and eggs, and sleep time from the model to check whether collinearity among some of the explanatory variables had any influence on the estimated coefficients. The results of the smaller model indicated that the original significance estimates changed only a little. Consumption of vegetables and fruits had a strong correlation of .51. However, we could not find significant association of vegetables with the response when replacing fruits with vegetables even in the smaller model.

Sensitivity Analysis

We believe that the random effects for modeling participant-to-participant variation and food intake among different people reduced bias of the estimates caused by unobserved factors of the participants. To further examine any selection bias resulting from the fact that the number of weekly entries per participant varied, we conducted one sensitivity analysis to check the robustness of our findings. We randomly selected at most three records per week from each diary keeper to form a subdataset for fitting the same mixed-effects model. The previously mentioned procedure was repeated 100 times to produce 100 sets of estimated coefficients. Then, we calculate the pooled estimate of each coefficient and standard errors of these pooled estimates (Multimedia Appendix 1). The significance levels of pooled estimates of these coefficients were mostly consistent with results of using the whole dataset. We found that only beans and pulses became marginally significant, and staying up late became insignificant.

We also tried another sensitivity analysis to include those participants filling in the diaries less than 10 days per month. By relaxing the inclusion criterion, we included an additional 42 participants in the model. The whole model was then rerun with 202 participants. The results (Multimedia Appendix 2) showed that significant levels of these variable estimates were still consistent with the main finding. The reason for choosing the inclusion criterion of at least 10 days per month (nearly 2-3 days per week) was to consider the short incubation period of influenza. If the participants regularly filled in the diaries, we still can capture the signals at either the early or late stage of an ILI episode.

Discussion

Principal Findings

The online diary-based approach in this study was used to collect not only ILI symptoms but also daily health behaviors and participants' social networks. This innovative approach can reduce recall bias compared with weekly or monthly surveys. Although the presence of significant risk factors for ILI has been revealed in different studies, few have quantitatively estimated the risk levels associated with different factors. In this study, we had the opportunity to collect all those risk factors from online diaries during an influenza season. We have used these empirical observations to reveal several risk reduction–related health behaviors such as avoiding contact with persons with ILI, sleeping earlier, keeping a good diet, exercising more, and being aware of environmental temperature and air pollution. Our study demonstrated a wide spectrum of ILI risks at the personal, contact, and environmental levels.

In traditional approaches to studying interpersonal influenza transmission, the main focus was identifying people with ILI symptoms or confirmed ILI in a hospital setting and then following up with the potential transmission within the household or in schools [10,30]. One study with a flu watch cohort in the United Kingdom also tried to recruit healthy participants from volunteers among general practitioners to monitor the influenza activity, severity, and virus evolution in the community [31]. The participants needed to record their contact patterns and activities before and during illness. However, these ILI symptoms were only collected from the ego. Without the ILI status of their contacted alters, we cannot observe influenza transmission. Similarly, in our study, we have self-reported ILI status of the diary keepers at least 2 to 3 days a week during the study period. ILI status of the alters was recorded only when they were contacted by the diary keepers, and this information was thus incomplete. The limited data create a hurdle for examining the dynamics of influenza transmission among alters in the egocentric networks although we can estimate the influence of alters on egos.

The household cohort can provide insights to capture the dynamics of influenza transmission after identifying an index case in a household [10]. It requires substantial resources to confirm the infected cases and closely follow up on the series of infections. The transmission setting focused on the household and family members. In this study, we tried to understand not only ILI transmission risk in the household but also in other settings where the participants had contact with others. In our online diary, the relationship between the participant and the person with whom they had contact and the place where they had contact were all recorded. Therefore, we did further stratification analysis for different relationships of contact. We stratified our contacts into two categories, contact with family members and contact with other people, to refit the mixed-effects model. The results showed that participants reporting contact with nonrelatives had higher risk of getting ILI than those who had contact with family members (Multimedia Appendix 3). It might be that participants knew of their family members' ILI symptoms early, so they were able to take some preventive measures. The other reason might be that more than 70% of our participants were college students and young adults who had high chances of close contact with classmates, good friends, and coworkers.

In this study, we were able to quantify the risk of developing ILI from having been in contact with persons with ILI in the past 3 days. The estimated OR of 1.87 in our model is quite similar to a household transmission study conducted in France, which also reported that the hazard ratio (HR) of increased risk of influenza transmission in preschool contacts was 1.85 compared with school age and adult contacts [32]. They also reported an increased risk in contacts exposed to preschool index patients (HR=1.93) and school-age index patients (HR=1.68), compared with those exposed to adult index cases.

We also found that having a later bedtime in the past 3 days is a significant risk factor for developing ILI. In children, staying up late has previously been found to be associated with poorer quality of life and overall health [33]. Although sleep duration and quality of sleep were not significantly associated with the chance of getting ILI in our study, later bedtime may serve as a proxy measure of sleeping quality. Good sleep is correlated with good immunity function, especially for the primary response to infection. One study in the United States recruited 153 healthy volunteers to observe the association between sleep duration and sleep efficiency and development of a clinical cold over 14 days [34]. Their results showed that participants with less than 7 hours of sleep were 2.94 times as likely (95% CI 1.18-7.30) to develop a cold than those with 8 or more hours of sleep. Poor sleep efficiency was also a very strong risk factor for developing a cold (OR 5.50, 95% CI 2.08-14.48) [34].

Diet and physical exercise are also very important health behaviors to enhance immunity and reduce the chance of influenza infection. Although we have considered many types of foods in the model, we finally found that fruits, beans and pulses, and dairy products are associated with lower ILI risk. In the literature, one study found that anthocyanins from fruit extracts inhibited influenza virus adsorption into cells and also virus release from infected cells [35]. One cohort study over 12 influenza seasons in Ontario, Canada, found that compared with inactive individuals, moderately active (OR 0.83, 95% CI 0.74-0.94) and active (OR 0.87; 95% CI 0.77-0.98) individuals were less likely to make an influenza-coded visit [36]. For those individuals aged less than 65 years, the protective effects of activity are still significant (active OR 0.86, 95% CI 0.75-0.98, moderately active: OR 0.85, 95% CI 0.74-0.97). Our result also showed that 30 min of exercise time was associated with a lower chance of getting ILI (OR 0.76, 95% CI 0.65-0.89).

In addition to personal risk factors, environmental factors such as temperature variation, $PM_{2.5}$, and O_3 exposure are also associated with immunity, virus transmission, and replication from the literature. In one guinea pig study, cold and dry conditions were best for influenza transmission between hosts [37]. Previous studies have focused more on how seasonal temperature variation affected influenza epidemics. One study showed that a 1°C decrease in temperature would increase the risk of influenza infection by 11% (OR 1.11, 95% CI 1.03-1.20) [38]. Although most studies did not find an effect of temperature variation on influenza activity, our results from this unique dataset imply that this relationship is worth further investigation.

Positive correlations between $PM_{2.5}$ and ILI have been found in many studies in Beijing, China [39,40]. They found that the effect on influenza incidence caused by $PM_{2.5}$ had an increasing gradient as $PM_{2.5}$ increased when $PM_{2.5}$ concentration was larger than 70 $\mu g/m^3$. Another study involving 47 Chinese cities also showed that the effect of ambient $PM_{2.5}$ on influenza incidence occurred at lag 2 day, with a relative risk of 1.015 (95% CI 1.004-1.025) associated with a 10$\mu g/m^3$ increase in $PM_{2.5}$ [15]. In addition to epidemiological findings showing a connection between ILI and $PM_{2.5}$, one study conducted in Taiwan isolated influenza virus and avian influenza virus from ambient samples during Asian dust storm days [41]. A possible mechanism is that influenza virus was carried by airborne particles and then inhaled into the respiratory tract. That study also found that ILI risk only increased with elevated $PM_{2.5}$ during flu season and was not significant during nonflu season. Another air pollutant, O_3, was also found to induce cleavage of the influenza A hemagglutinin (HA) protein, and the secreted endogenous proteases were sufficient to activate HA, leading to a significant increase in viral replication [16].

One social network study [12] found that people with more social ties might have higher chances of getting infection or spreading one to others. In our study, we also included the average number of contact persons in the past 3 days in the model, but the results did not show statistical significance. One nationwide survey in Taiwan found that people with a higher number of contacts also had a higher likelihood of getting an influenza vaccination [42]. Thus, people with more social ties will also take some preventative measures against influenza; that might be the reason we could not find a correlation between ILI and the number of social ties in this study.

Limitations

There are several limitations in this study. The first one is a lack of laboratory samples to confirm whether the participants or the contact alters were in fact influenza infected or not. From a practical viewpoint, it is not possible to know who transmitted the influenza virus among the general public. Thus, we treat this uncertainty as a random effect for each participant. Furthermore, we used a relatively strict definition of ILI in our questionnaire. We treated participants as getting ILI only when they selected “definitely having ILI symptoms” and clicked specific symptoms in the symptom list. Using this definition, our data showed a good match with the epi-curve of ILI from the nationwide outpatient surveillance. The second limitation is the sample size and representativeness. Because of the nature of our study design, the diary-based follow-up needs patience and persistence to record the health and contact diary for 6 months. It is difficult to keep a large number of participants in a study of this kind for such a long time. Also, volunteer-based online surveys can never claim to be representative of the general population. Due to the limitation of small sample size, we cannot generalize our findings to the general population. In particular, we have found that the participants in this study tended to be young, and there was a large proportion of female participants (76.3%, 122/160). Therefore, we included age and sex in the model for adjustment. In addition, we also incorporated other variables such as geographic area into the model, with no significant difference.

The third limitation is lack of data on protective behaviors such as wearing a facial mask and vaccination status during the study period. In fact, we collected the vaccination status before and after influenza season. However, we did not know the exact date of vaccination. Therefore, it was not available to measure the effect of this diary-based study design. The fourth limitation is lack of exercise intensity in this diary. We only asked participants to record the duration of exercise. From the current findings, the exercise time should be longer than 30 min to show reduced risk of getting ILI. The fifth limitation is related to exposure estimation. In fact, we did not know exactly where the participants were located. They only reported the township where they lived. Therefore, we were only able to use a station in the corresponding township or the closest station to it. The concentrations of $PM_{2.5}$ and ozone could be treated only as environmental exposure, not as a measure of personal exposure.

Conclusions

In conclusion, our study shows that keeping a healthier lifestyle, including having a nutritious diet, sleeping earlier, and doing longer physical exercise is associated with a lower risk of getting ILI. Self-protection and avoiding contact with infected persons, as well as keeping alert to temperature changes and air quality are also linked to lower risk of getting ILI.

Acknowledgments

This research was supported by a grant from the Academia Sinica, Taiwan, R.O.C. (AS-103-TP-C03). The authors would like to thank Ms Jie-Yu Sung for the administrative support.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Comparison of estimated coefficients using the whole dataset and pooled estimates of 100 sampled subdatasets.

[PDF File (Adobe PDF File), 25KB - publichealth_v4i2e40_app1.pdf]

Multimedia Appendix 2

Sensitivity analysis on model estimation by including those participants filling in the diaries less than 10 days per month.

[PDF File (Adobe PDF File), 34KB - publichealth_v4i2e40_app2.pdf]

Multimedia Appendix 3

Estimation of ILI risk by stratifying two types of contacts with infected family members and nonrelatives.

[PDF File (Adobe PDF File), 34KB - publichealth_v4i2e40_app3.pdf]

References

1. de Francisco Shapovalova N, Donadel M, Jit M, Hutubessy R. A systematic review of the social and economic burden of influenza in low- and middle-income countries. Vaccine 2015 Nov 27;33(48):6537-6544. [doi: 10.1016/j.vaccine.2015.10.066] [Medline: 26597032]
2. Nair H, Brooks WA, Katz M, Roca A, Berkley JA, Madhi SA, et al. Global burden of respiratory infections due to seasonal influenza in young children: a systematic review and meta-analysis. Lancet 2011 Dec 03;378(9807):1917-1930. [doi: 10.1016/S0140-6736(11)61051-9] [Medline: 22078723]
3. Reed C, Chaves SS, Daily Kirley P, Emerson R, Aragon D, Hancock EB, et al. Estimating influenza disease burden from population-based surveillance data in the United States. PLoS One 2015 Mar 04;10(3):e0118369 [FREE Full text] [doi: 10.1371/journal.pone.0118369] [Medline: 25738736]
4. Chung JR, Flannery B, Thompson MG, Gaglani M, Jackson ML, Monto AS, et al. Seasonal effectiveness of live attenuated and inactivated influenza vaccine. Pediatrics 2016 Feb;137(2):e20153279 [FREE Full text] [doi: 10.1542/peds.2015-3279] [Medline: 26738884]
5. Hirve S, Lambach P, Paget J, Vandemaele K, Fitzner J, Zhang W. Seasonal influenza vaccine policy, use and effectiveness in the tropics and subtropics - a systematic literature review. Influenza Other Respir Viruses 2016 Jul;10(4):254-267 [FREE Full text] [doi: 10.1111/irv.12374] [Medline: 26842617]
6. Smith SM, Sonego S, Wallen GR, Waterer G, Cheng AC, Thompson P. Use of non-pharmaceutical interventions to reduce the transmission of influenza in adults: A systematic review. Respirology 2015 Aug;20(6):896-903 [FREE Full text] [doi: 10.1111/resp.12541] [Medline: 25873071]

7. DeStefano F, Haber M, Currivan D, Farris T, Burrus B, Stone-Wiggins B, et al. Factors associated with social contacts in four communities during the 2007-2008 influenza season. Epidemiol Infect 2011 Aug;139(8):1181-1190. [doi: 10.1017/S095026881000230X] [Medline: 20943002]
8. Ibuka Y, Ohkusa Y, Sugawara T, Chapman GB, Yamin D, Atkins KE, et al. Social contacts, vaccination decisions and influenza in Japan. J Epidemiol Community Health 2016 Feb;70(2):162-167 [FREE Full text] [doi: 10.1136/jech-2015-205777] [Medline: 26424846]
9. Cauchemez S, Bhattarai A, Marchbanks TL, Fagan RP, Ostroff S, Ferguson NM, Pennsylvania H1N1 working group. Role of social networks in shaping disease transmission during a community outbreak of 2009 H1N1 pandemic influenza. Proc Natl Acad Sci U S A 2011 Feb 15;108(7):2825-2830 [FREE Full text] [doi: 10.1073/pnas.1008895108] [Medline: 21282645]
10. Tsang TK, Lau LL, Cauchemez S, Cowling BJ. Household transmission of influenza virus. Trends Microbiol 2016 Feb;24(2):123-133 [FREE Full text] [doi: 10.1016/j.tim.2015.10.012] [Medline: 26612500]
11. Wu S, Ma C, Yang Z, Yang P, Chu Y, Zhang H, et al. Hygiene behaviors associated with influenza-like illness among adults in Beijing, China: a large, population-based survey. PLoS One 2016 Feb 03;11(2):e0148448 [FREE Full text] [doi: 10.1371/journal.pone.0148448] [Medline: 26840614]
12. Christakis NA, Fowler JH. Social network sensors for early detection of contagious outbreaks. PLoS One 2010 Sep 15;5(9):e12948 [FREE Full text] [doi: 10.1371/journal.pone.0012948] [Medline: 20856792]
13. Roussel M, Pontier D, Cohen JM, Lina B, Fouchet D. Quantifying the role of weather on seasonal influenza. BMC Public Health 2016 May 26;16:441 [FREE Full text] [doi: 10.1186/s12889-016-3114-x] [Medline: 27230111]
14. Lowen AC, Mubareka S, Steel J, Palese P. Influenza virus transmission is dependent on relative humidity and temperature. PLoS Pathog 2007 Oct 19;3(10):1470-1476 [FREE Full text] [doi: 10.1371/journal.ppat.0030151] [Medline: 17953482]
15. Chen G, Zhang W, Li S, Zhang Y, Williams G, Huxley R, et al. The impact of ambient fine particles on influenza transmission and the modification effects of temperature in China: a multi-city study. Environ Int 2017 Jan;98:82-88. [doi: 10.1016/j.envint.2016.10.004] [Medline: 27745688]
16. Kesic MJ, Meyer M, Bauer R, Jaspers I. Exposure to ozone modulates human airway protease/antiprotease balance contributing to increased influenza A infection. PLoS One 2012;7(4):e35108 [FREE Full text] [doi: 10.1371/journal.pone.0035108] [Medline: 22496898]
17. Chan TC, Yen TJ, Fu YC, Hwang JS. ClickDiary: online tracking of health behaviors and mood. J Med Internet Res 2015 Jun 15;17(6):e147 [FREE Full text] [doi: 10.2196/jmir.4315] [Medline: 26076583]
18. Dalton C, Carlson S, Butler M, Cassano D, Clarke S, Fejsa J, et al. Insights from flutracking: thirteen tips to growing a web-based participatory surveillance system. JMIR Public Health Surveill 2017 Aug 17;3(3):e48 [FREE Full text] [doi: 10.2196/publichealth.7333] [Medline: 28818817]
19. Baltrusaitis K, Santillana M, Crawley AW, Chunara R, Smolinski M, Brownstein JS. Determinants of participants' follow-up and characterization of representativeness in Flu Near You, a participatory disease surveillance system. JMIR Public Health Surveill 2017 Apr 07;3(2):e18 [FREE Full text] [doi: 10.2196/publichealth.7304] [Medline: 28389417]
20. Click Diary. 2017. URL: http://clickdiary.stat.sinica.edu.tw/ [accessed 2017-08-29] [WebCite Cache ID 6t4hoXYbw]
21. Wen CP, Wai JP, Tsai MK, Yang YC, Cheng TY, Lee MC, et al. Minimum amount of physical activity for reduced mortality and extended life expectancy: a prospective cohort study. Lancet 2011 Oct 01;378(9798):1244-1253. [doi: 10.1016/S0140-6736(11)60749-6] [Medline: 21846575]
22. Read JM, Lessler J, Riley S, Wang S, Tan LJ, Kwok KO, et al. Social mixing patterns in rural and urban areas of southern China. Proc Biol Sci 2014 Apr 30;281(1785):20140268 [FREE Full text] [doi: 10.1098/rspb.2014.0268] [Medline: 24789897]
23. Chan TC, Fu YC, Hwang JS. Changing social contact patterns under tropical weather conditions relevant for the spread of infectious diseases. Epidemiol Infect 2015 Jan;143(2):440-451. [doi: 10.1017/S0950268814000843] [Medline: 24725605]
24. Gosling SD, Rentfrow PJ, Swann Jr. WB. A very brief measure of the Big-Five personality domains. J Res Pers 2003 Dec;37(6):504-528. [doi: 10.1016/S0092-6566(03)00046-1]
25. Chang YH, Fu YC. The evolution of the Taiwan Social Change Survey. In: Birth of JGSS and Its Fruit: Social Surveys in Different Countries and Areas and JGSS. Presented at: JGSS Symposium 2003; 2003; East Osaka City p. 149-160.
26. Newbold P, Carlson W, Thorne B. Statistics for Business and Economics, 7th Edition. London, UK: Pearson; 2009.
27. Taiwan Environmental Protection Administration. 2017. Air quality monitoring network URL: http://taqm.epa.gov.tw/taqm/tw/YearlyDataDownload.aspx [accessed 2017-08-29] [WebCite Cache ID 6t4iLYoRo]
28. R Core Team. R-project. Vienna, Austria: R Foundation for Statistical Computing; 2016. R: A language and environment for statistical computing URL: https://www.r-project.org/ [accessed 2018-03-08] [WebCite Cache ID 6xmPOmxHD]
29. Venables WN, Ripley BD. Modern Applied Statistics With S. New York: Springer Science+Business Media; 2002.
30. Kamigaki T, Mimura S, Takahashi Y, Oshitani H. Analysis of influenza transmission in the households of primary and junior high school students during the 2012-13 influenza season in Odate, Japan. BMC Infect Dis 2015 Jul 23;15:282 [FREE Full text] [doi: 10.1186/s12879-015-1007-8] [Medline: 26201316]
31. Fragaszy EB, Warren-Gash C, Wang L, Copas A, Dukes O, Edmunds WJ, on behalf of the Flu Watch Group. Cohort Profile: The Flu Watch Study. Int J Epidemiol 2017 Apr 01;46(2):e18. [doi: 10.1093/ije/dyv370] [Medline: 26940466]
32. Viboud C, Boëlle PY, Cauchemez S, Lavenu A, Valleron AJ, Flahault A, et al. Risk factors of influenza transmission in households. Br J Gen Pract 2004 Sep;54(506):684-689 [FREE Full text] [Medline: 15353055]

33. Chen X, Sekine M, Hamanishi S, Wang H, Gaina A, Yamagami T, et al. Lifestyles and health-related quality of life in Japanese school children: a cross-sectional study. Prev Med 2005 Jun;40(6):668-678. [doi: 10.1016/j.ypmed.2004.09.034] [Medline: 15850863]
34. Cohen S, Doyle WJ, Alper CM, Janicki-Deverts D, Turner RB. Sleep habits and susceptibility to the common cold. Arch Intern Med 2009 Jan 12;169(1):62-67 [FREE Full text] [doi: 10.1001/archinternmed.2008.505] [Medline: 19139325]
35. Knox YM, Hayashi K, Suzutani T, Ogasawara M, Yoshida I, Shiina R, et al. Activity of anthocyanins from fruit extract of Ribes nigrum L. against influenza A and B viruses. Acta Virol 2001;45(4):209-215. [Medline: 11885927]
36. Siu E, Campitelli MA, Kwong JC. Physical activity and influenza-coded outpatient visits, a population-based cohort study. PLoS One 2012;7(6):e39518 [FREE Full text] [doi: 10.1371/journal.pone.0039518] [Medline: 22737242]
37. Lowen AC, Steel J. Roles of humidity and temperature in shaping influenza seasonality. J Virol 2014 Jul;88(14):7692-7695. [doi: 10.1128/JVI.03544-13] [Medline: 24789791]
38. Jaakkola K, Saukkoriipi A, Jokelainen J, Juvonen R, Kauppila J, Vainio O, KIAS-Study Group. Decline in temperature and humidity increases the occurrence of influenza in cold climate. Environ Health 2014 Mar 28;13(1):22 [FREE Full text] [doi: 10.1186/1476-069X-13-22] [Medline: 24678699]
39. Feng C, Li J, Sun W, Zhang Y, Wang Q. Impact of ambient fine particulate matter (PM2.5) exposure on the risk of influenza-like-illness: a time-series analysis in Beijing, China. Environ Health 2016 Feb 11;15:17 [FREE Full text] [doi: 10.1186/s12940-016-0115-2] [Medline: 26864833]
40. Liang Y, Fang L, Pan H, Zhang K, Kan H, Brook JR, et al. PM2.5 in Beijing – temporal pattern and its association with influenza. Environ Health 2014 Dec 4;13:102. [doi: 10.1186/1476-069X-13-102] [Medline: 25471661]
41. Chen PS, Tsai FT, Lin CK, Yang CY, Chan CC, Young CY, et al. Ambient influenza and avian influenza virus during dust storm days and background days. Environ Health Perspect 2010 Sep;118(9):1211-1216 [FREE Full text] [doi: 10.1289/ehp.0901782] [Medline: 20435545]
42. Chan TC, Fu YC, Wang DW, Chuang JH. Determinants of receiving the pandemic (H1N1) 2009 vaccine and intention to receive the seasonal influenza vaccine in Taiwan. PLoS One 2014 Jun 27;9(6):e101083 [FREE Full text] [doi: 10.1371/journal.pone.0101083] [Medline: 24971941]

Abbreviations

ILI: influenza-like illness
HR: hazard ratio
IQR: interquartile range
OR: odds ratio
PM2.5: particulate matter

Edited by G Eysenbach; submitted 31.08.17; peer-reviewed by J Kwong, Y Ibuka, J Chua; comments to author 29.11.17; revised version received 26.01.18; accepted 11.02.18; published 09.04.18.

Please cite as:
Chan TC, Hu TH, Hwang JS
Estimating the Risk of Influenza-Like Illness Transmission Through Social Contacts: Web-Based Participatory Cohort Study
JMIR Public Health Surveill 2018;4(2):e40
URL: http://publichealth.jmir.org/2018/2/e40/
doi:10.2196/publichealth.8874
PMID:29631987

Original Paper

Accurately Inferring Compliance to Five Major Food Guidelines Through Simplified Surveys: Applying Data Mining to the UK National Diet and Nutrition Survey

Nicholas Rosso[1]; Philippe Giabbanelli[2], PhD

[1]Data Analytics for Complex Human Behaviors Laboratory, Computer Science Department, Northern Illinois University, DeKalb, IL, United States
[2]Data Analytics for Complex Human Behaviors Laboratory, Department of Computer Science, Furman University, Greenville, SC, United States

Corresponding Author:
Philippe Giabbanelli, PhD
Data Analytics for Complex Human Behaviors Laboratory
Department of Computer Science
Furman University
3300 Poinsett Hwy
Greenville, SC, 29613
United States
Phone: 1 864 294 2097
Email: giabbanelli@gmail.com

Abstract

Background: National surveys in public health nutrition commonly record the weight of every food consumed by an individual. However, if the goal is to identify whether individuals are in compliance with the 5 main national nutritional guidelines (sodium, saturated fats, sugars, fruit and vegetables, and fats), much less information may be needed. A previous study showed that tracking only 2.89% of all foods (113/3911) was sufficient to accurately identify compliance. Further reducing the data needs could lower participation burden, thus decreasing the costs for monitoring national compliance with key guidelines.

Objective: This study aimed to assess whether national public health nutrition surveys can be further simplified by only recording whether a food was consumed, rather than having to weigh it.

Methods: Our dataset came from a generalized sample of inhabitants in the United Kingdom, more specifically from the National Diet and Nutrition Survey 2008-2012. After simplifying food consumptions to a binary value (1 if an individual consumed a food and 0 otherwise), we built and optimized decision trees to find whether the foods could accurately predict compliance with the major 5 nutritional guidelines.

Results: When using decision trees of a similar size to previous studies (ie, involving as many foods), we were able to correctly infer compliance for the 5 guidelines with an average accuracy of 80.1%. This is an average increase of 2.5 percentage points over a previous study, showing that further simplifying the surveys can actually yield more robust estimates. When we allowed the new decision trees to use slightly more foods than in previous studies, we were able to optimize the performance with an average increase of 3.1 percentage points.

Conclusions: Although one may expect a further simplification of surveys to decrease accuracy, our study found that public health dietary surveys can be simplified (from accurately weighing items to simply checking whether they were consumed) while improving accuracy. One possibility is that the simplification reduced noise and made it easier for patterns to emerge. Using simplified surveys will allow to monitor public health nutrition in a more cost-effective manner and possibly decrease the number of errors as participation burden is reduced.

(JMIR Public Health Surveill 2018;4(2):e56) doi:10.2196/publichealth.9536

KEYWORDS

diet, food, and nutrition; public health informatics; supervised machine learning

Introduction

Insufficient compliance with dietary guidelines can lead to several health problems, whereas following guidelines can have protective effects. Systematic reviews have linked excess salt consumption with increased blood pressure, which raises the risk for cardiovascular diseases [1,2]. Furthermore, other meta-reviews have found that a higher consumption of fruit and vegetables "was significantly associated with a lower risk of all-cause mortality" [3]. It is, thus, essential to monitor compliance with such guidelines to understand and improve a population's health. To assess whether guidelines are followed, data on nutritional intake must be compiled. A comprehensive assessment of nutritional intake can be burdening (as individuals need to record the exact amount and type of foods consumed), which may in part cause the inaccuracies found when individuals provide such reports [4].

Data mining is a computational technique (often equated with machine learning), which offers significant potential to alleviate that burden by finding key patterns in data. Among the different tasks performed in data mining, our focus is on *classification*, which consists of automatically relating a set of feature variables (eg, age, gender, food consumed) to an outcome (eg, being in compliance with guidelines on salt). Classification has been increasingly used in recent years for research on several weight-related outcomes, such as obesity [5-7], nutrition [8], and physical activity [9].Classification has demonstrated its potential to complement statistical regressions, particularly for nonlinear phenomena (as is often the case with human behaviors [10] such as eating behaviors), and without requiring a priori assumptions on the relationship between patterns and outcomes [11]. In particular, classification has been applied on several occasions to find the key questions that surveys need to infer a target behavior. For instance, in the case of adolescent binge drinking, researchers showed that rules in a household were strongly linked with the outcome, whereas other dimensions (eg, communication) were not as salient [12]. Similarly, previous research in public health nutrition found that only 2.89% (113/3911) of the food items were required to infer compliance to the 5 major national guidelines [8].

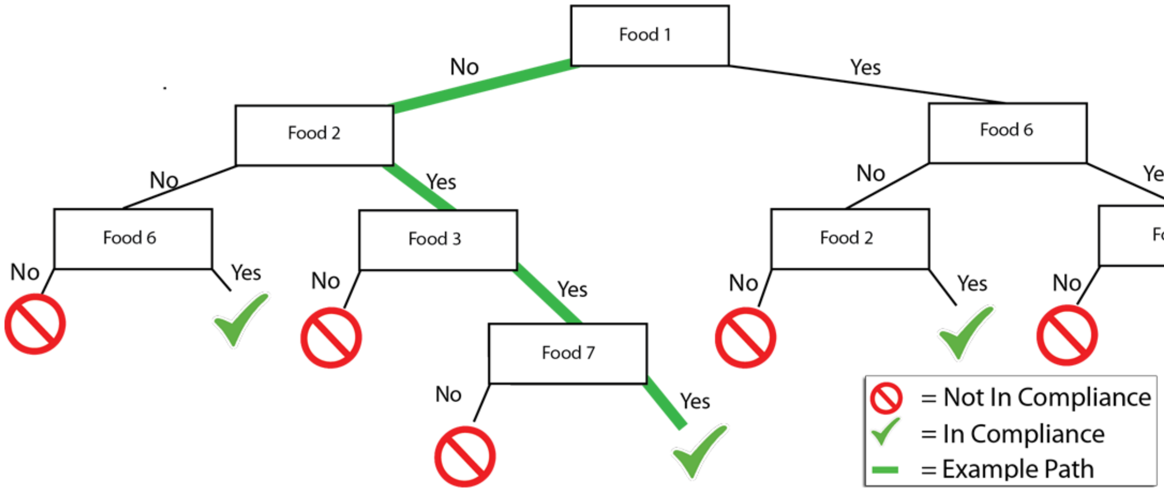
There are many algorithms to choose from when performing classification. Decision trees in particular have proven to be a popular approach [6-9,11-13] for at least 2 reasons. First, they can then be used as a visual tool: instead of being a black-box model (such as a deep neural network or a support vector machine), they clearly articulate the rules that transform the description of a new participant's case into an outcome (Figure 1). Second, these rules can also be used as flowcharts in public health, or clinical settings, to support decision-making activities (eg, triage) [14,15]. In line with these studies, this paper employs the classification technique of decision trees.

Our overarching goal is to further simplify public health nutrition surveys. Building on previous work showing that only 2.89% (113/3911) of the items were necessary [8] to infer compliance with major food guidelines, we will assess whether survey items can be reduced to binaries (was a food eaten or not?) rather than requiring an accurate weight. To identify *success* in adequately simplifying surveys, we will compute whether decision trees can still accurately infer compliance with guidelines using the simplified surveys. Specifically, we will simplify items in the National Diet and Nutrition Survey (NDNS) 2008-2012 to binary and assess whether decision trees built on the simplified dataset are about as accurate as decision trees built on the initial dataset.

The principal contributions of this study can be summarized as follows:

- We demonstrate that simplifying the information recorded in a specific dietary survey is not necessarily detrimental to identifying key public health outcomes.
- The application of our work to dietary public health suggests that nutritional surveys may be simplified when the aim is to predict compliance with major nutritional guidelines. This simplification may reduce participation burden, lower study costs, or increase the sample size at a same cost.
- The methodological part of our work illustrates the potential for data mining to contribute to public health not only by making predictions, but by identifying what part of the data is truly needed to form these predictions.

Figure 1. A decision tree starts at a root (top). For a given individual, we repeatedly compare the individual’s data with the questions in the tree. In this example, if the individual did not consume food 1, then the follow-up question is whether food 2 was consumed. Eventually, we reach a conclusion: whether the individual was in compliance with the guideline or not. Such trees are automatically built from the data.



Methods

Data Used

Our dataset came from a generalized sample of inhabitants from the United Kingdom: the National Diet and Nutrition Survey (NDNS) 2008-2012. The NDNS data were obtained from the UK Data Archive [16]. The NDNS is a cross-sectional survey that records the nutrient intake as well as the nutritional status of the population within the United Kingdom [17]. To allow for comparison with previous studies [8], we used data from years 1-4 of this program, collected in 2008-2012. The NDNS collected data from a sample of 1000 respondents per year, consisting of adults and children aged 18 months and above. Households across the United Kingdom were selected to take part in the NDNS using a multistage probability design. During each wave, a random sample of primary sampling units was selected for inclusion. These are small geographical areas that allow for more efficient data collection by being geographically focused.

Within the dataset, food consumption at a daily level is recorded for participants over several days. To record portion sizes, common household measures (eg, one tablespoon, one cup) and weight in grams were used for the foods consumed throughout the study, including the consumption of liquids. Foods are described specifically and can be related to other foods in a subgroup or a group. For instance, the consumption of bananas would be entered as with 3 different levels of detail: as individual foods (eg, bananas raw flesh only), as subfood groups (eg, bananas), or as food groups (eg, fruit and vegetables).

The NDNS dataset only contains the foods consumed, their composition, and demographical information. It does not make any conclusion in regard to nutritional guidelines. The dataset was expanded in a previous study to include this information [8]. This was realized via the following process: (1) compute how much each individual consumed with respect to the 5 key dietary guidelines, then (2) compare this consumption with nutritional recommendations (which may be age-dependent), and (3) record the result as “Yes” when the participant was in compliance for a specific guideline or “No” when the participant was not in compliance. The detailed process is as follows.

The NDNS dataset has 4156 participants including 1189 children younger than 11 years. First, for each of the 4156 participants, compute the mean daily intake of fruit and vegetables and sodium, as well as the main daily percentage of energy derived from fat, saturated fat, and free sugars. Then, compare each individual's numbers with the corresponding nutritional recommendations to determine whether the individual is in compliance with the recommendation. UK recommendations on *fruit and vegetables* apply only to those aged 11 years or older, thus 1189 participants were excluded for this specific comparison. To be recorded as “Yes,” those retained needed to consume at least five 80-g portions of fruit and vegetable daily, allowing for at most 1 portion of juice. Although UK recommendations on *sodium* are also dependent on the age category, they adjust the comparison rather than excluding participants. A participant would be labeled as “Yes” if the sodium intake does not exceed [18] 2400 mg/d for those aged 11 years and older, 2000 mg/d for those aged 7-10 years, 1200 mg/d for those aged 4-6 years, and 800 mg/d for those aged 1-3 years.

The World Health Organization (WHO) recommends limitations on how much energy can be derived from each of the following categories: at most 30% from fat, at most 10% from saturated fat, and at most 10% from free sugars (sixth table in [19]). We then computed how much energy a participant derived from each category. If the energy derived from fat, saturated fat, and free sugars were under the WHO threshold, then we set the corresponding guideline to “YES.”

For each participant, our final dataset includes selected data from the NDNS survey (age, gender, and consumption for all of the 3911 individual foods) and additional data computed through the process above (whether or not they were in compliance for each of the 5 nutritional guidelines).

Methods Employed: Classification Using Decision Trees

A *classifier* is a model automatically built from a subset of the data (called *training set*) in which we know both the predictor variables (ie, age, gender, and foods eaten) and the class outcomes (ie, whether or not each of the 5 guidelines was met). The intention is to build "good" classifiers, that is, models that learn and generalize from the training set so that they can accurately predict the outcomes when presented with new cases [20]. Numerous methods build classifiers, such as support vector machines, decision trees, and rulesets [21]. As detailed in the Introduction, our study uses decision trees, which are a commonly used approach [6-9,11-13] that provides a usable visual tool (Figure 1) to support decision-making activities such as triage.

There are 2 types of classifications: binary and multi-class. In a binary situation, the outcome we seek to predict can only have 2 different values. Conversely, in a multi-classification problem, the outcome has 3 or more values. Our study focuses on a binary classification problem: for each of the 5 guidelines, we want to know whether or not the guideline is met.

The process to create a decision tree for binary classification has been detailed in numerous reference material such as Maimon and Rokach [22]; thus, we provide only a brief overview of this process. The dataset (detailed in the previous section) comes in as a spreadsheet, where rows correspond to individuals and columns represent their features (ie, their age, gender, diet, and whether or not each of the 5 guidelines was met). The goal is to train a decision tree so it automatically identifies the combination of predictor variables (age, gender, individual foods) to determine the class outcome (for each of the five guidelines). A small portion of the rows are used as the *training set* to guide the decision tree algorithms to produce specific trees. The algorithm will repeatedly subdivide the data, where the variable used to subdivide is represented as a node in the tree (Figure 1), and the subdivisions corresponding to different values are shown as branches leaving this node. For instance, Figure 1 shows that the first division is based on the hypothetical "food 1": one subdivision is produced when the food was not consumed (left branch), and the other subdivision corresponds to consuming this food (right branch).

A portion of the data is not provided to the algorithm for building the tree and is instead held to evaluate the quality of the generated tree [20]. This portion is called the *testing set*. To avoid basing our evaluation from one specific portion of the data that may not be representative, a process known as *cross-fold validation* divides the dataset into multiple portions, building the tree on one (training) and evaluating it on the others (testing) before repeating the division until all parts have been used for training and testing. This common process to evaluate classification accuracy helps prevent overfitting, where performances on the training set are very good but its generalization on the training set performs poorly [20]. The evaluation consists of presenting the tree with individuals from the testing set and asking what the classes should be. Then, the tree predicts a class outcome, which we compare with the real outcome from the dataset. The extent to which these outcomes match is called the *accuracy*. When the outcomes are binary, the percentage of "Yes" instances correctly classified is known as *recall*, and the percentage of "No" instances correctly classified is known as *specificity*. Intuitively, accuracy is the performance of the model across class outcomes, whereas recall and specificity are performances for one outcome in particular.

Highlighting recall and sensitivity is useful when the costs of making mistakes may be different: in health studies, giving someone an intervention that they do not need may be a very different issue from initially suggesting not to give them the intervention that they need. In addition, datasets are frequently imbalanced, that is, there can be many more cases for one outcome than the other. In this case, a high accuracy may be misleading as the tree may do well for the most common case, while being very inaccurate for the less common case. By providing the recall and sensitivity, our study supports public health officials in evaluating our performance by giving more or less weight to specific outcomes. As in previous work, our overall accuracy assumes that the error costs are similar [8], that is, concluding that someone does not follow guidelines while they do is no worse than concluding that they follow guidelines while they do not. Assuming different error costs would need additional evidence, and it would also lead to different methods as relatively few approaches can mine data under differential error costs [23,24].

In general, class imbalance can be addressed by eliminating cases of the majority class (undersampling), creating new cases for the minority class (oversampling), or biasing the classification algorithm (eg, using nonuniform error costs on the classes) [12,24]. For this study, we use sampling techniques. Specifically, we used Synthetic Minority Over-Sampling Technique, or SMOTE for short. As concluded by Batista et al, "over-sampling methods in general, and SMOTE-based methods in particular" were very efficient to address class imbalances [5,25].Although a comprehensive discussion on class balancing is beyond the scope of this study, we note that finding good approaches for synthetic over-sampling remains a very active area of research, as even popular methods such as SMOTE have weaknesses. However, such weaknesses are particularly encountered when dealing with very high-dimensional datasets such as text [26], which is not the case here.

Overall Process

Our process is summarized in Figure 2. We start with the same dataset as used in our previous study: the NDNS 2008-2012 data expanded with compliance to each guideline [8].We departed from the previous study [8] by simplifying the dataset: we only recorded whether an item was consumed (1) or not (0). These data are given as input to the classification process, which was performed 5 times, for each of the guidelines. For a given guideline, we removed the compliance of the 4 other guidelines from the dataset. We do not want the algorithm to use compliance on fat to infer compliance on saturated fat: instead, compliance should be inferred from the foods, age, and gender only. As discussed in the previous subsection, balancing needs to be performed to avoid biasing the algorithm in favor of the most common outcome. We used SMOTE to ensure that both outcomes (meeting or not meeting a guideline) occur with the same prevalence. The balanced dataset was then fed into the Weka software version 3.7, maintained by the Machine Learning Group at the University of Waikato. We used the J48 decision tree algorithm, which implements the highly cited C4.5 algorithm by Ross Quinlan [27].

Like most classification algorithms, C4.5 (and its J48 implementations) take parameters that can impose further constraints on the resulting tree. We tested different parameter values to either (1) find the most accurate decision tree with a similar structure (ie, number of foods) to the trees generated in the previous study using the exact weights of foods, or (2) identify the most accurate tree without consideration for the number of foods involved [8]. These allow to perform two operations. First, we can *compare* with the previous study [8], in which the tree built for each guideline used a very small number of foods. Our objective was to constrain our new tree in using a similar number of foods, such that we can observe how accuracy changes when foods are simplified (in this study) instead of being recorded exactly (in the previous study). To lower the number of foods used by the algorithm, we increased the minimum number of cases required to further cut the data (ie, add a decision node to the tree). Second, we seek to *optimize*, by identifying how accurate we can be using our simplified foods, possibly at the expense of using more foods.

After each tree was built, we used 10-fold cross-validation [10,20,28]. This method for evaluation divides the dataset into 10 equal parts. Nine parts were used for the training set, and one for the testing set. After the process was repeated 10 times, the evaluation was conducted on all of the data, and the average results were reported. For full disclosure, all of our decision trees are available on the Open Science Framework platform (see [29]).

A sample of our approach to explore the trade-off between the number of foods and accuracy is illustrated in Table 1, showing a parameter sweep by increasing the minimum number of instances to (nonmonotonically) reduce the number of foods used. The rationale for this process is as follows. For the decision tree algorithm to create a new branch, it needs to find where to “cut” in the dataset. If there are not enough instances to cut, then a new branch will not be made. When this new branch *would* have been based on a factor not previously used in the tree, then preventing its creation limits the number of foods used. However, the branch may have involved an already existing factor. Raising the minimum number of instances thus limits opportunities for the algorithm to involve additional factors. Table 1 exemplifies how the number of factors tends to decrease as the minimum number of instances increases.

In the guiding example of Table 1, we predict adherence to the guideline on free sugars. The previous study used 28 foods for this class [8], thus we seek the highest accuracy that we can achieve with 28 foods or less. The best trade-off is found using a minimum of 95 instances, leading to 25 foods and an accuracy of 77.9%, which is higher than the 76.5% previously found. This trade-off would thus be reported in our results.

Table 1 exemplifies our methodology on choosing a decision tree comparable with the previous study [8], by changing the minimum number of instances. We observe that, as this number increases, the number of factors tends to decrease. The goal is to find the result with the highest accuracy while using no more foods than in the previous study.

Figure 2. Flow diagram of our methodology, showing the acquisition, preprocessing, and mining of the data. NDNS: National Diet and Nutrition Survey; SMOTE: Synthetic Minority Over-Sampling Technique.

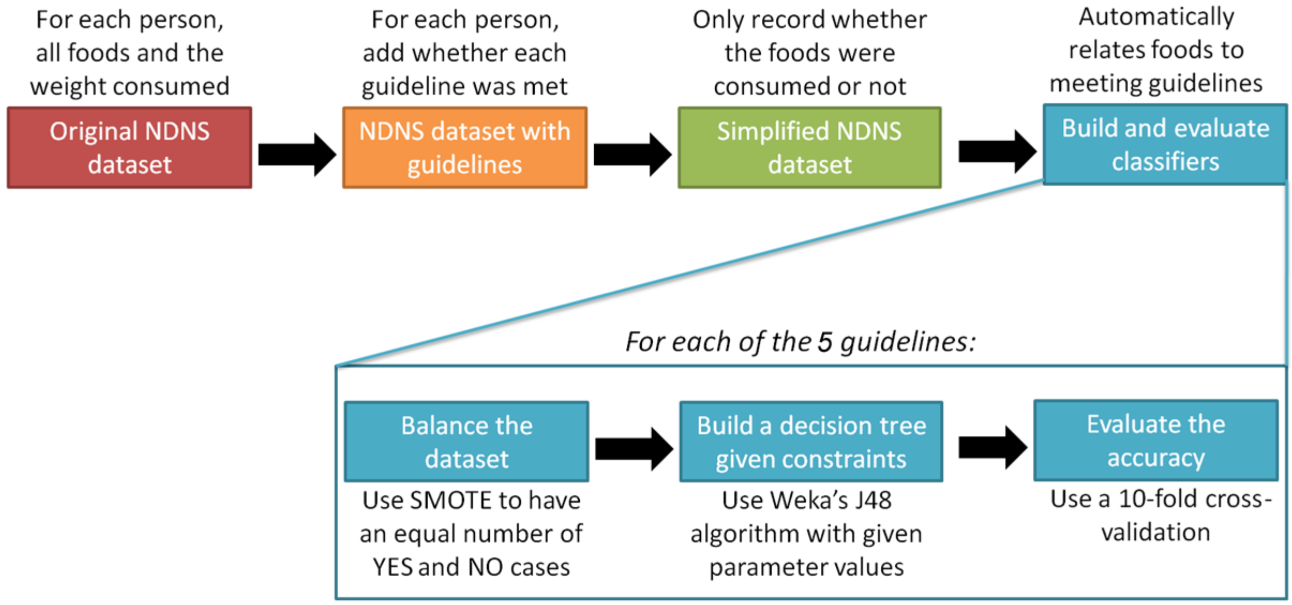


Table 1. Sample outcome for the decision tree classifier on free sugars.

Study	Minimum number of instances	Accuracy (average)	Recall	Specificity	Number of factors
Previous	60	76.5	76.1	76.9	28
Current	60	78.2	73.6	82.9	31
Current	70	78.1	74.7	77.3	31
Current	80	78.3	74.7	78.3	30
Current	90	77.9	75.1	80.8	30
Current	95	77.9	75.1	80.7	25
Current	100	77.3	75.7	78.9	26
Current	115	77.2	75.5	78.8	22

Results

Our dataset can broadly be understood as consisting of participants (the rows) and their food consumptions (the columns). Demographic characteristics of the participants (regardless of food consumptions) are summarized in Table 2 including gender, nationality, marital status, and economic status. Participants were on average 30.5 (SD 20.9) years old. Patterns of food consumption are shown in Figure 3. As will be shown in our results, it is not because a food is common that it should be included to identify whether participants meet a dietary guideline.

The methods introduced in the previous section select a food if it helps to separate individuals in compliance versus those who are not. For instance, if eating bananas is highly prevalent in the population, then knowing whether a person ate bananas may not be useful to predict dietary compliance. Conversely, if a food was clearly associated with a healthier diet for a handful of individuals, the frequency may be too low to warrant its inclusion at the population level.

Our new decision trees, built on simplified reporting of foods, were slightly more accurate than previous trees built using the exact weighted foods. This was found across all guidelines (Table 3). In 4 out of the 5 guidelines, the increase in accuracy was particularly noticeable to infer that someone did not meet a guideline. For instance, the previously reported accuracy of 78.4% [8] on finding noncompliance with fat had now increased to 88.5%. The increase in finding noncompliant cases was met in 2 guidelines (salt, free sugar) with a small decrease in accuracy for compliant cases, whereas it was similar in a third guideline (fat).

Across the 5 guidelines, our new decision trees had an accuracy of 80.1%. That is, in 4 out of 5 cases, by only knowing whether foods were consumed, and using at most a few dozen foods, we can successfully conclude whether nutritional guidelines are met. This accuracy is 2.6 percentage points higher than the average on previous decision trees (77.5%). That is, not asking individuals to weigh foods leads to being better able to tell if they meet guidelines.

The optimized classifiers performed slightly better with an average accuracy of 80.6% on classified classes (Table 4). The optimized trees also had an average percentage increase of 3.1 points from the previous classifiers see (Figure 4). In all guidelines but one (salt), the increase in performance was obtained at the expense of using more foods. Although the number of foods used can increase by up to 50% (for saturated fat, fruits and vegetables), the absolute number of foods remains very small compared with the initial NDNS data and its 3911 foods.

To better contrast optimized decision trees versus those limited in the number of foods, Figure 4 shows where they led to either better (green) or lower (red) accuracy compared with the previous study [8]. Both methods generally underperformed on finding noncompliance to fruit and vegetables, and on finding compliance on salt and free sugars. They over-performed on fat and saturated fat. In summary, the consequences of simplifying dietary surveys are not uniform across guidelines, as some will see a small reduction in accuracy, whereas others may see a large improvement, resulting in the average accuracy (across all guidelines) being improved.

In Table 5, we list all individual foods used at least 5 times in predicting compliance with the guidelines, using either decision trees similar to the previous study [8], or the optimized trees. The expanded list of foods used one or more times is provided as supplementary material online [29]. Note that foods used to predict compliance with a guideline may not be part of what counts within this guideline. For instance, sausage rolls are neither fruit nor vegetables, yet they are used to predict fruit and vegetables consumption. We also observe that these foods are not necessarily the “common” ones shown in Figure 3.

Table 2. Key characteristics of the National Diet and Nutrition Survey (NDNS) household dataset. All participants in the study were within the United Kingdom. There were several study waves, with around 1000 respondents per year.

Characteristics	Categorical count, n (%)
Gender	
Male	5034 (47.41)
Female	5439 (52.57)
Within compliance	
Free sugars	1472 (35.41)
Salt	2524 (60.73)
Fat	1045 (25.14)
Saturated fat	795 (19.13)
Fruits and vegetables	656 (15.78)
Nationality	
English	5036 (48.08)
Northern Irish	3442 (32.86)
Scottish	684 (6.53)
Welsh	398 (3.80)
Irish	194 (1.85)
Other	719 (6.88)
Marital status	
Single (never married)	6240 (59.57)
Married (living with partner)	1960 (18.71)
Divorced	261 (2.49)
Married (living separate)	3 (0.06)
Widowed	139 (1.32)
Other	1870 (17.85)
Economic status	
Going to school full-time	2974 (28.39)
Full or part time employment	4440 (42.39)
Not working presently	3039 (29.02)

Figure 3. Main foods either by (a) contribution to caloric intake, or (b) prevalence among individuals.

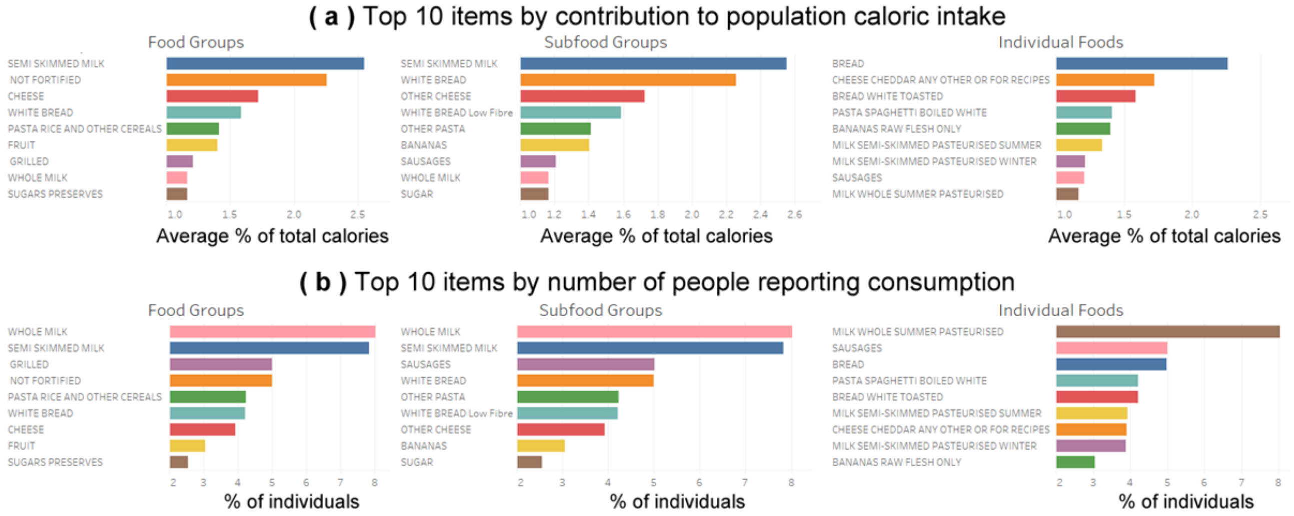


Table 3. Comparison of the best decision tree using the weight of foods (previous study, Giabbanelli and Adams, 2016 [8]) or simplified foods (this study), while keeping the number of foods similar.

Study	Guidelines	Number of instances	Accuracy (%)	Recall	Specificity	Number of factors
Previous	Free sugars	60	76.5	76.1	76.9	28
Current	Free sugars	95	77.9	75.1	80.7	25
Previous	Fat	70	72.4	66.3	78.4	33
Current	Fat	90	79.4	70.4	88.5	33
Previous	Fruits and vegetables	50	83.1	82.5	83.8	11
Current	Fruits and vegetables	90	82.2	82.3	82.2	10
Previous	Saturated fat	20	79.7	75.8	83.6	28
Current	Saturated fat	90	84.6	77.4	91.8	27
Previous	Salt	15	75.8	81.9	69.8	28
Current	Salt	55	76.3	79.5	73.2	26

Table 4. Comparison of the best decision tree using the weight of foods (previous study, Giabbanelli and Adams, 2016 [8]) or simplified foods (this study), without being limited by the number of foods.

Study	Guidelines	Number of instances	Accuracy (%)	Recall	Specificity	Number of factors
Previous	Free sugars	60	76.5	76.1	76.9	28
Current	Free sugars	60	78.2	73.6	82.9	31
Previous	Fat	70	72.4	66.3	78.4	33
Current	Fat	70	79.9	72.3	87.7	43
Previous	Fruits and vegetables	50	83.1	82.5	83.8	11
Current	Fruits and vegetables	50	83.5	84.9	82.2	16
Previous	Saturated fat	20	79.7	75.8	83.6	28
Current	Saturated fat	20	84.7	79.3	90.1	42
Previous	Salt	15	75.8	81.9	69.8	28
Current	Salt	50	76.6	79.9	73.2	25

Figure 4. Accuracy, recall (“Yes”), and specificity (“No”) when (a) limiting the number of foods as in a previous study (Giabbanelli & Adams, 2016 [8]), or (b) using any number of foods to build the decision trees, giving us the optimized decision trees.

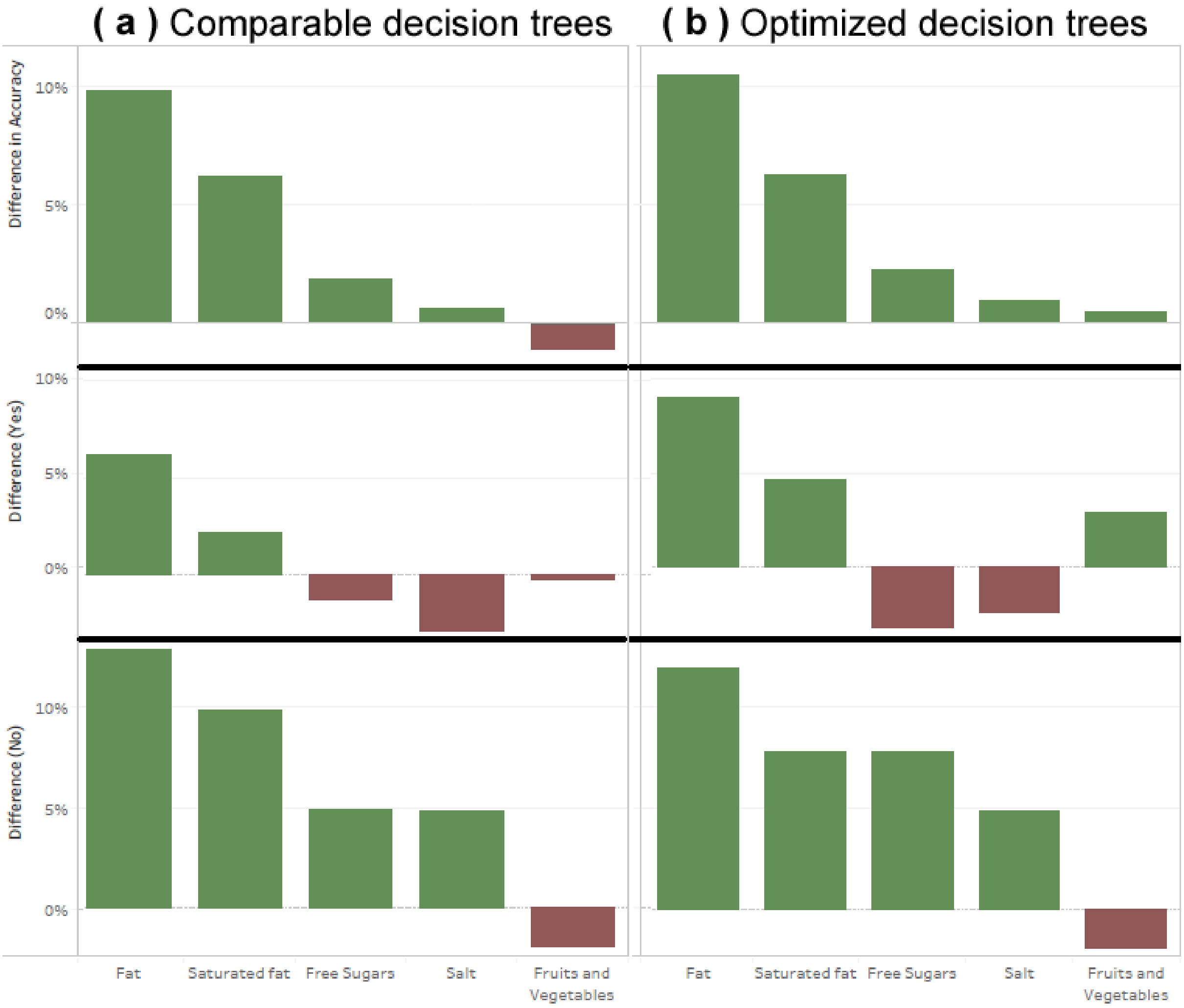


Table 5. Individual foods used as predictors at least 5 times in the trees generated using our 2 processes (similar/optimized) and for the 5 guidelines: Fruit and Vegetables, Fat, Saturated Fat, Salt, and Free Sugars. The frequency is the number of times that a food is used as a decision node across all trees (eg, if used 3 times in 5 trees each, it would be 15).

Variables	Similar decision tree					Optimized decision tree					Total frequency
	FV[b]	Fat	SatFat[c]	Salt	Sug[d]	FV	Fat	SatFat	Salt	Sug	
Individual food											
Sausages		✓	✓	✓			✓	✓	✓		20
Bananas raw	✓	✓			✓	✓	✓	✓			19
Sausage roll	✓	✓	✓			✓	✓	✓			16
Cheese cheddar		✓	✓	✓			✓	✓	✓		14
Milk chocolate		✓	✓				✓	✓			12
Butter salted		✓		✓		✓		✓		✓	10
Cheese spreads		✓						✓			8
Ice cream		✓						✓		✓	8
Fruit drink					✓			✓		✓	8
Chicken pieces		✓					✓				8
Sex		✓		✓					✓		7
Potato crisps		✓		✓							6
Apples	✓					✓					6
Milk whole		✓	✓					✓			6
Beans baked		✓		✓			✓		✓		6
Onions	✓			✓							6
Cola					✓					✓	6
Apple juice unsweetened UHT[a]	✓				✓	✓				✓	6
Olive oil	✓										6
Orange juice unsweetened					✓					✓	6
Orange juice unsweetened UHT										✓	6
Bacon				✓							6
Apple juice unsweetened						✓				✓	5
Demographic											
Sex		✓		✓					✓		7

[a] UHT: Ultra-high-temperature processing.

[b] FV: fruits and vegetables.

[c] SatFat: saturated fat.

[d] Sug: free sugars.

Discussion

Principal Findings

Monitoring at the national level whether the population is in compliance with an array of nutritional guidelines currently requires an extensive data collection process, in which individuals report and weigh the exact foods that they consumed. Our previous study demonstrated that only 2.89% (113/3911) of the foods needed to be reported to predict with 77.5% accuracy (72%-83% across guidelines) whether individuals achieve key dietary recommendations regarding sodium, saturated fats, sugars, fruit/vegetables, and fats [8]. In this study, we investigated the consequences of further simplifying reporting by only asking participants whether they ate a specific food rather than having to weight it.

Although we may have expected a decreased accuracy as a consequence of removing information, our results paradoxically indicate that accuracy has improved to 80%. We observed that results were particularly improved when inferring compliance to the guidelines on fat and saturated fat, but a trade-off was operated on free sugars and salt where a decrease in recall was counter-balanced by a larger increase in specificity. Results were more nuanced on fruit and vegetables, where optimized decision trees were able to offset a loss of specificity with a higher gain in recall (thus resulting in higher accuracy), but nonoptimized decision trees resulted in a small loss of accuracy. Overall, these findings suggest that foods may not have to be weighted, but this may depend on (1) which food guidelines need to be monitored and (2) whether public health officials decide that recall is more important than sensitivity (or vice versa) instead of giving them equal weight.

The main applications of our results are twofold. First, we may simplify surveys not only by asking for few foods in adaptive questionnaires (as shown in [8]) but also by asking binary questions "Did you consume this food?" rather than requiring participants to provide an exact weight. This contribution will result in more time-effective assessments and may lower the cognitive effort required from participants, which in turn can decrease the error rate. Second, identifying a few questions yielding an accuracy of 80% is most applicable when a trade-off has to be found between accuracy and participation burden. For instance, a doctor may have many tools and physiological measures as part of the treatment process (eg, blood pressure, HbA1c), and including a few dietary questions with an accuracy of 80% may be more feasible than a more thorough survey. For population health, our work is particularly applicable in large studies where only a limited number of questions can be used to investigate a *subgroup within arms*. For instance, in the Netherlands, the nationwide Longitudinal Internet Studies for the Social sciences (LISS) panel sends questionnaires each month, dealing with many topics ranging from alcohol [30] to happiness. Nutrition would only be one part, and a reduced measurement approach would be necessary.

Comparison With Other Dietary Methods

There are several alternatives to the analysis conducted here. First, an index-based analysis consists of a scoring system based on a priori knowledge that researchers have about (1) dietary guidance and (2) the scores to assign for sets of dietary components based on the guidance. This analysis can be used to assess adherence to guidelines [31-33] or summarize an individual's diet quality [2,31]. Within epidemiology, indices are used to identify the risk an individual will have to certain diseases based a combination of foods [31]. Although the reliance of indices on a priori knowledge makes them less sensitive to variations in the sample than our method, they may (depending on their design and structure) require more foods and accuracy in portion sizes. Considering the trade-off or "continuum" from few simple questions to favoring high accuracy, indices can lead to a higher accuracy than the method presented here but may not be as amenable to a "reduced" form as a short addendum to a large panel study such as LISS [30]. We also note that the transparency and simplicity of decision trees can support practitioners in interpreting the rules (eg, for triage) with little to no training, whereas dietary indices can produce summary scores where expertise is still important for interpretation.

Second, one could perform a cluster analysis. As summarized by Reedy et al, "clusters are driven by the sample from which they are derived, so their applicability as a standard for evaluating diets of different populations is limited because of the number of factors that determine food selection" [34]. Cluster analysis is an *unsupervised* data mining technique that identifies similarities between groups based on their patterns of food consumption: for instance, "fatty meats" may be an important similarity between men [34]. This is different from the classification approach taken here, which is a *supervised* data mining technique that seeks to predict an outcome.

Finally, Food Frequency Questionnaires (FFQs) can provide a cost-effective approach to monitoring the health of a large population. Molag et al [35], as well as Noethlings et al, suggested that portion sizes may not be necessary [36]: "We conclude that the omission of individual portion size information would probably result in a notable reduction of interindividual variance. However, to reduce the respondents' burden and to increase data completeness in self-administration in large epidemiologic studies, the assignment of a constant portion size seems to be adequate." Our study confirms this finding while pointing out that accuracy may even increase; however, the effect depends on which guideline we monitor.

Strengths and Limitations

Our study aimed to determine the effects of reducing the level of details employed by a national dietary survey. The NDNS survey used here has been the subject of many publications and provides a wealth of high-quality data. However, several limitations stem from using this survey. First, the NDNS survey relies on self-reported food intake. Individuals may consciously, or unconsciously, misreport their consumption within a 24-hour time frame [4,37,38]. Using the exact weigh of foods is thus sensitive to misreporting, which was a limitation of our previous study [8]. In contrast, this study is not sensitive to misreporting how much of a food was consumed: it only takes into account whether *any* consumption of this food occurred. Reporting errors affecting our study would thus be to entirely ignore a specific food that was consumed or to report a food that was not consumed.

Second, this survey was specific to the population of the United Kingdom, as can be seen in the specific foods used as predictors. This limitation of the data entails that our conclusion may not be generalized to populations that have important differences in eating behaviors. In this case, our approach can be replicated by collecting the complete dataset (in the first study wave) and then using data mining to investigate the consequences of simplifying it (for future study waves). Replicating results across target populations is necessary before concluding that monitoring compliance to nutritional guidelines may generally be simplified.

Our study used the data mining technique of decision trees to automatically relate individual food consumption to meeting specific guidelines. This is a well-researched technique, which has been applied to problems arising in health on multiple occasions. One specific advantage of decision trees lies in their ability to produce a model that can easily be interpreted and used with limited training. For instance, in triage, decision trees provide a "flowchart" that lay participants as well as field specialists can use intuitively. That is, an adaptive questionnaire can be formed by following the rules induced by a tree (Figure 1), which can be done using a computer program or by individuals. In contrast, many other techniques (eg, Support

Vector Machines, Neural Networks) produce “black box” models, which are meant to be executed by machines rather than being read by humans. Future studies primarily concerned with accuracy (rather than transparency/readability of the model) may explore using such techniques. Contrasting the use of neural networks to decision trees over the same dataset would provide valuable insight on how accurate we can be without restrictions, which would help to better situate the results from this study.

Conclusions

We sought to determine whether identifying individual dietary compliance can be further simplified while remaining as informative and accurate. We found that reporting very few foods and only whether they were consumed was sufficient to correctly identify compliance to 5 major nutritional guidelines. Being able to reduce the detail of a dataset for national monitoring can make it easier to increasing monitoring frequency or monitor more participants, thus increasing research participations without increasing study costs.

Acknowledgments

The authors wish to thank the Department of Computer Science and the College of Liberal Arts and Sciences at Northern Illinois University for funding this study via start-up funds to PG. NR was also financially supported by the Office of Student Engagement and Experiential Learning at Northern Illinois University.

Conflicts of Interest

None declared.

References

1. Trieu K, McLean R, Johnson C, Santos JA, Raj TS, Campbell NR, et al. The Science of Salt: A Regularly Updated Systematic Review of the Implementation of Salt Reduction Interventions (November 2015 to February 2016). J Clin Hypertens (Greenwich) 2016 Dec;18(12):1194-1204 [FREE Full text] [doi: 10.1111/jch.12909] [Medline: 27704719]
2. Owen AJ, Retegan C, Rockell M, Jennings G, Reid CM. Inertia or inaction? Blood pressure management and cardiovascular risk in diabetes. Clin Exp Pharmacol Physiol 2009 Jul;36(7):643-647. [doi: 10.1111/j.1440-1681.2008.05125.x] [Medline: 19076166]
3. Wang X, Ouyang Y, Liu J, Zhu M, Zhao G, Bao W, et al. Fruit and vegetable consumption and mortality from all causes, cardiovascular disease, and cancer: systematic review and dose-response meta-analysis of prospective cohort studies. BMJ 2014;349:g4490.
4. Poslusna K, Ruprich J, de Vries JH, Jakubikova M, van't Veer P. Misreporting of energy and micronutrient intake estimated by food records and 24 hour recalls, control and adjustment methods in practice. Br J Nutr 2009 Jul;101(Suppl 2):S73-S85. [doi: 10.1017/S0007114509990602] [Medline: 19594967]
5. Batista G, Prati R, Monard M. Balancing Strategies and Class Overlapping. In: Famili AF, Kok JN, Peña JM, Siebes A, Feelders A, editors. Advances in Intelligent Data Analysis VI. Lecture Notes in Computer Science. Berlin, Heidelberg: Springer; 2005:24-35.
6. Nau C, Ellis H, Huang H, Schwartz BS, Hirsch A, Bailey-Davis L, et al. Exploring the forest instead of the trees: An innovative method for defining obesogenic and obesoprotective environments. Health Place 2015 Sep;35:136-146 [FREE Full text] [doi: 10.1016/j.healthplace.2015.08.002] [Medline: 26398219]
7. Seyednasrollah F, Mäkelä J, Pitkänen N, Juonala M, Hutri-Kähönen N, Lehtimäki T, et al. Prediction of Adulthood Obesity Using Genetic and Childhood Clinical Risk Factors in the Cardiovascular Risk in Young Finns Study. Circ Cardiovasc Genet 2017 Jun;10(3):e001554. [doi: 10.1161/CIRCGENETICS.116.001554] [Medline: 28620069]
8. Giabbanelli PJ, Adams J. Identifying small groups of foods that can predict achievement of key dietary recommendations: data mining of the UK National Diet and Nutrition Survey, 2008-12. Public Health Nutr 2016 Jun;19(9):1543-1551 [FREE Full text] [doi: 10.1017/S1368980016000185] [Medline: 26879185]
9. Kerr J, Patterson RE, Ellis K, Godbole S, Johnson E, Lanckriet G, et al. Objective Assessment of Physical Activity: Classifiers for Public Health. Med Sci Sports Exerc 2016 May;48(5):951-957 [FREE Full text] [doi: 10.1249/MSS.0000000000000841] [Medline: 27089222]
10. Huys R, Jirsa VK(. Nonlinear Dynamics in Human Behavior. Studies in Computational Intelligence 328. Berlin Heidelberg: Springer-Verlag; 2011.
11. Crutzen R, Giabbanelli P. Using Classifiers to Identify Binge Drinkers Based on Drinking Motives. Subst Use Misuse 2015;49(1-2):110. [doi: 10.3109/10826084.2013.824467] [Medline: 23964957]
12. Garcia VL, Sánchez JS, Mollineda RA, Alejo R, Sotoca JM. The class imbalance problem in pattern classification and learning. 2009 Presented at: the Congreso Español de Informática (CEDI); 2007; Zaragoza, Spain.
13. Crutzen R, Giabbanelli PJ, Jander A, Mercken L, de Vries H. Identifying binge drinkers based on parenting dimensions and alcohol-specific parenting practices: building classifiers on adolescent-parent paired data. BMC Public Health 2015 Aug 05;15:747 [FREE Full text] [doi: 10.1186/s12889-015-2092-8] [Medline: 26243154]

14. Tanner L, Schreiber M, Low JG, Ong A, Tolfvenstam T, Lai Y, et al. Decision tree algorithms predict the diagnosis and outcome of dengue fever in the early phase of illness. PLoS Negl Trop Dis 2008 Mar 12;2(3):e196 [FREE Full text] [doi: 10.1371/journal.pntd.0000196] [Medline: 18335069]
15. Steele R, Green SM, Gill M, Coba V, Oh B. Clinical decision rules for secondary trauma triage: predictors of emergency operative management. Ann Emerg Med 2006 Feb;47(2):135. [doi: 10.1016/j.annemergmed.2005.10.018] [Medline: 16431223]
16. Gov.UK. 2014 May. NDNS: results from Years 1 to 4 (combined) URL: https://www.gov.uk/government/statistics/national-diet-and-nutrition-survey-results-from-years-1-to-4-combined-of-the-rolling-programme-for-2008-and-2009-to-2011-and-2012 [accessed 2017-09-25] [WebCite Cache ID 6tk7fopKD]
17. Webarchive.nationalarchives.gov.uk. 2010. National Diet and Nutrition Survey: Headline Results from Year 1 of the Rolling Programme and comparison with previous surveys London: Food Standards Agency and Department of Health URL: http://webarchive.nationalarchives.gov.uk/20101210143146/http://www.food.gov.uk/multimedia/pdfs/publication/ndnsreport0809.pdf [accessed 2018-05-13] [WebCite Cache ID 6zOfNs9aU]
18. Scientific Advisory Committee on Nutrition. Salt and Health. London, UK: The Stationery Office; 2003.
19. Apps.WHO. 2003. Diet, Nutrition and the Prevention of Chronic Diseases. Report of a Joint WHO/FAO Expert Consultation. WHO Technical Report Series no 916 URL: http://apps.who.int/iris/bitstream/10665/42665/1/WHO_TRS_916.pdf [accessed 2018-05-04] [WebCite Cache ID 6zAVQNyTn]
20. Bramer M. Estimating the Predictive Accuracy of a Classifier. In: Principles of Data Mining. Undergraduate Topics in Computer Science. London: Springer; 2013:79-92.
21. Giabbanelli PJ, Peters JG. An Algebraic Approach to Combining Classifiers. Procedia Comput Sci 2015;51:1545-1554.
22. Rokach L, Maimon O. In: Bunke H, Wang PSP, editors. Data Mining with Decision Trees:Theory and Applications. 2nd Edition. Machine Perception and Artificial Intelligence Series. Vol 81. Singapore: World Scientific Publishing Co. Pte. Ltd; Oct 2014.
23. Weiss G, McCarthy K, Zabar B. Cost-Sensitive Learning vs Sampling: Which is Best for Handling Unbalanced Classes with Unequal Error Costs. 2007 Presented at: Proceedings of the International conference on Data Mining; 2007; Washington DC, USA.
24. Maloof MA. Learning When Data Sets are Imbalanced and When Costs are Unequal and Unknown. 2003 Presented at: International Conference of Machine Learning; 2003; Washington DC, USA.
25. Chawla NV, Bowyer KW, Hall LO, Kegelmeyer WP. SMOTE: Synthetic Minority Over-sampling Technique. J Artif Intell Res 2002 Jun 01;16:321-357.
26. Bellinger C. Ruor.uottawa. 2016. Beyond the Boundaries of SMOTE: A Framework for Manifold-Based Synthetically Oversampling URL: https://ruor.uottawa.ca/handle/10393/34643 [accessed 2018-05-04] [WebCite Cache ID 6zAVjiALQ]
27. Ross Quinlan J. C4.5: programs for machine learning. San Francisco: Morgan Kaufmann; 1993:1993.
28. Kohavi R. A study of cross-validation and bootstrap for accuracy estimation and model selection. 1995 Presented at: 14th international joint conference on Artificial Intelligence; August 20 - 25, 1995; Montreal QC, Canada p. 1137-1143.
29. Osf.io. OSF: Data Mining on Foods URL: https://osf.io/znv82/ [accessed 2018-05-11] [WebCite Cache ID 6zKonPcvj]
30. Giabbanelli PJ, Crutzen R. An agent-based social network model of binge drinking among Dutch adults. J Artif Soc Soc Simul 2013;16(2):10.
31. Kant AK. Indexes of overall diet quality: a review. J Am Diet Assoc 1996 Aug;96(8):785-791. [doi: 10.1016/S0002-8223(96)00217-9] [Medline: 8683010]
32. Hu FB. Dietary pattern analysis: a new direction in nutritional epidemiology. Curr Opin Lipidol 2002 Feb;13(1):3-9. [Medline: 11790957]
33. Schwingshackl L, Hoffmann G. Diet quality as assessed by the Healthy Eating Index, the Alternate Healthy Eating Index, the Dietary Approaches to Stop Hypertension score, and health outcomes: a systematic review and meta-analysis of cohort studies. J Acad Nutr Diet 2015 May;115(5):780-800.e5. [doi: 10.1016/j.jand.2014.12.009] [Medline: 25680825]
34. Reedy J, Wirfält E, Flood A, Mitrou PN, Krebs-Smith SM, Kipnis V, et al. Comparing 3 dietary pattern methods--cluster analysis, factor analysis, and index analysis--With colorectal cancer risk: The NIH-AARP Diet and Health Study. Am J Epidemiol 2010 Feb 15;171(4):479-487 [FREE Full text] [doi: 10.1093/aje/kwp393] [Medline: 20026579]
35. Molag ML, de Vries JH, Ocké MC, Dagnelie PC, van den Brandt PA, Jansen MC, et al. Design characteristics of food frequency questionnaires in relation to their validity. Am J Epidemiol 2007 Dec 15;166(12):1468-1478. [doi: 10.1093/aje/kwm236] [Medline: 17881382]
36. Noethlings U, Hoffmann K, Bergmann MM, Boeing H, European Investigation into Cancer and Nutrition. Portion size adds limited information on variance in food intake of participants in the EPIC-Potsdam study. J Nutr 2003 Feb;133(2):510-515. [doi: 10.1093/jn/133.2.510] [Medline: 12566492]
37. Wehling H, Lusher J. People with a body mass index 30 under-report their dietary intake: A systematic review. J Health Psychol 2017 Jul 01:1359105317714318. [doi: 10.1177/1359105317714318] [Medline: 28810493]
38. Gibson RS, Charrondiere UR, Bell W. Measurement Errors in Dietary Assessment Using Self-Reported 24-Hour Recalls in Low-Income Countries and Strategies for Their Prevention. Adv Nutr 2017 Nov 15;8(6):980-991. [doi: 10.3945/an.117.016980] [Medline: 29141979]

Abbreviations

LISS: Longitudinal Internet Studies for the Social sciences
NDNS: National Diet and Nutrition Survey
SMOTE: Synthetic Minority Over-Sampling Technique
WHO: World Health Organization

Edited by T Sanchez; submitted 30.11.17; peer-reviewed by A Selya, R Beheshti, P Wark; comments to author 12.01.18; revised version received 07.03.18; accepted 13.04.18; published 30.05.18.

Please cite as:
Rosso N, Giabbanelli P
Accurately Inferring Compliance to Five Major Food Guidelines Through Simplified Surveys: Applying Data Mining to the UK National Diet and Nutrition Survey
JMIR Public Health Surveill 2018;4(2):e56
URL: http://publichealth.jmir.org/2018/2/e56/
doi:10.2196/publichealth.9536
PMID:

Corrigenda and Addenda

Correction: Predicting Sexual Behaviors Among Homeless Young Adults: Ecological Momentary Assessment Study

Diane Santa Maria[1], DrPH, RN, MSN; Nikhil Padhye[1], PhD; Yijiong Yang[2], MA; Kathryn Gallardo[2], MSW; Michael Businelle[3], PhD

[1]School of Nursing, University of Texas Health Science Center at Houston, Houston, TX, United States
[2]School of Public Health, University of Texas Health Science Center at Houston, Houston, TX, United States
[3]Department of Family and Preventive Medicine, The University of Oklahoma Health Sciences Center, Oklahoma City, OK, United States

Corresponding Author:
Diane Santa Maria, DrPH, RN, MSN
School of Nursing
University of Texas Health Science Center at Houston
6901 Bertner Ave
Houston, TX, 77030
United States
Phone: 1 713 500 2187
Fax: 1 713 500 2142
Email: diane.m.santamaria@uth.tmc.edu

Related Article:

Correction of: http://publichealth.jmir.org/2018/2/e39/

(JMIR Public Health Surveill 2018;4(2):e10806) doi:10.2196/10806

The authors of the paper "Predicting Sexual Behaviors Among Homeless Young Adults: Ecological Momentary Assessment Study" (JMIR Public Health Surveill 2018;4(2):e39) made an error in the reporting of odds ratios (ORs) and the labeling of the columns in Table 4. While it does not change the interpretation of the data or outcomes, the values need to be corrected.

In the Abstract, the OR values for same-day drug use and sexual urge in the sentence "The estimated odds ratios (ORs) were notable for same-day drug use (OR 2.17, 95% CI 4.48-17.31; *P*<.001) and sexual urge (OR 1.44, 95% CI 1.60-11.28; *P*=.004)" have been changed to 8.80 and 4.23, respectively.

In the "Risk Estimator for Sexual Activity Days" subsection of the Results, the OR values for same-day drug use and sexual urge have again been changed to 8.80 and 4.23, respectively. The subsequent text which previously read "The odds of having sex on a given day is more than double for every unit increase in drug use. The odds of having sex rise by 44% for every unit increase in sexual urge" has been changed to the following:

> The odds of having sex increase 8.8 times on days when youth use drugs after adjusting for sexual urge and other predictors. The odds of having sex increase 4.2 times per unit increase in sexual urge, after adjusting for drug use and other predictors.

In Table 4, column 2 was mistakenly labeled "OR" and is now "Coefficient B". Column 4 was mistakenly labeled "Exp(B)" and is now "OR". The data values for these columns have not been changed. The correct header for the table appears below.

The corrected article will appear in the online version of the paper on the JMIR website on May 7, 2018, together with the publication of this correction notice. Because this was made after submission to PubMed, Pubmed Central, and other full-text repositories, the corrected article also has been re-submitted to those repositories.

Table 4.

Variable	Coefficient B	SE	OR	Z	*P* value	95% CI of OR

Edited by G Eysenbach; submitted 17.04.18; this is a non–peer-reviewed article;accepted 17.04.18; published 07.05.18.

Please cite as:
Santa Maria D, Padhye N, Yang Y, Gallardo K, Businelle M
Correction: Predicting Sexual Behaviors Among Homeless Young Adults: Ecological Momentary Assessment Study
JMIR Public Health Surveill 2018;4(2):e10806
URL: http://publichealth.jmir.org/2018/2/e10806/
doi:10.2196/10806
PMID:30578237

Corrigenda and Addenda

Figure Caption Correction: Characteristics of Articles About Human Papillomavirus Vaccination in Japanese Newspapers: Time-Series Analysis Study

Nao Ueda[1], BA; Ryoki Yokouchi[2], BA; Taro Onoda[3], BS (Med); Atsushi Ogihara[4], PhD

[1]Graduate School of Human Sciences, Waseda University, Tokorozawa, Japan
[2]Department of Health Sciences and Social Welfare, School of Human Sciences, Waseda University, Tokorozawa, Japan
[3]Department of Medicine, School of Medicine, Tokai University, Isehara, Japan
[4]Faculty of Human Sciences, Waseda University, Tokorozawwa, Japan

Corresponding Author:
Nao Ueda, BA
Graduate School of Human Sciences
Waseda University
2-579-15
Mikajima
Tokorozawa, 359-1192
Japan
Phone: 81 80 5082 3970
Email: naaaaa0428@gmail.com

Related Article:

Correction of: http://publichealth.jmir.org/2017/4/e97/

(JMIR Public Health Surveill 2018;4(2):e10878) doi:10.2196/10878

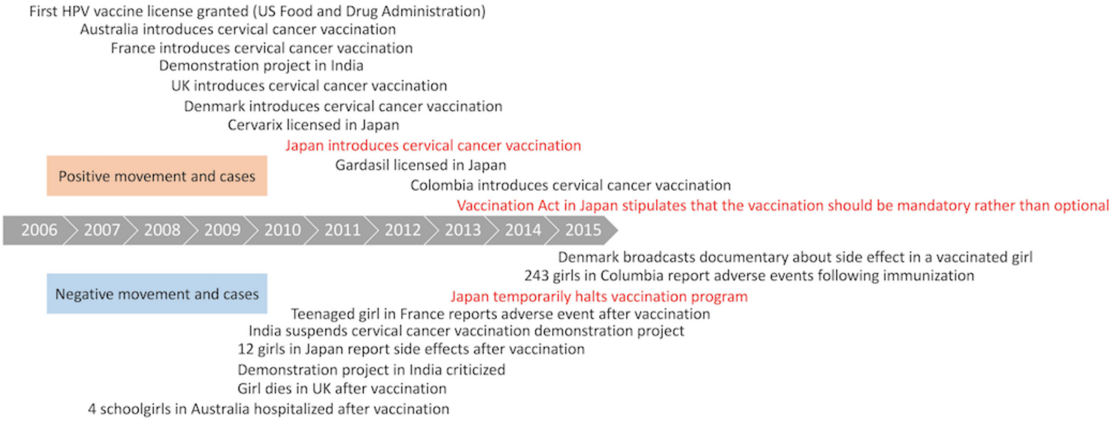
The authors of the paper “Characteristics of Articles About Human Papillomavirus Vaccination in Japanese Newspapers: Time-Series Analysis Study” (JMIR Public Health Surveill 2017;3(4):e97) omitted information in the caption of Figure 2. It should have read “Worldwide trends in cervical cancer vaccination. HPV: human papillomavirus. (Modified by the authors based on Wilson et al [6]).”

The corrected caption appeared in the online version of the paper on the JMIR website on March 16, 2018, together with the publication of a correction notice (J Med Internet Res 2018;20(3):e27). Because this was made after submission to PubMed or Pubmed Central and other full-text repositories, the corrected article was also re-submitted to those repositories.

However, the original correction notice published on March 16, 2018, erroneously appeared in a different journal (J Med Internet Res) than the original paper (JMIR Public Health and Surveillance), and was retracted on May 10, 2018, together with republication of this correction in the correct journal.

Please see the figure with the corrected caption here.

Figure 2. Worldwide trends in cervical cancer vaccination. HPV: human papillomavirus. (Modified by The authors based on Wilson et al).



Edited by G Eysenbach; submitted 25.04.18; this is a non–peer-reviewed article;accepted 25.04.18; published 10.05.18.

Please cite as:
Ueda N, Yokouchi R, Onoda T, Ogihara A
Figure Caption Correction: Characteristics of Articles About Human Papillomavirus Vaccination in Japanese Newspapers: Time-Series Analysis Study
JMIR Public Health Surveill 2018;4(2):e10878
URL: http://publichealth.jmir.org/2018/2/e10878/
doi:10.2196/10878
PMID:30578211

Corrigenda and Addenda

Expression of Editorial Concern, Correction of Conflict of Interest and Affiliation, and Data Corrections

JMIR Editorial Office

Related Articles:

Correction of: http://publichealth.jmir.org/2018/1/e6/

Correction of: https://publichealth.jmir.org/2019/3/e16314

(JMIR Public Health Surveill 2018;4(2):e53) doi:10.2196/publichealth.9932

Expression of Editorial Concern

We are hereby expressing serious concerns over conflicts of interests (COI) by the authors of a JMIR Research Protocols paper [1] as well as a subsequent JMIR Public Health & Surveillance [2] paper (where the results of the protocol were published). While we think that complementary medicine (as well as associated lifestyle changes, technologies promoting such lifestyle changes, and other tools and behavioral innovations covered by JMIR journals) deserve to be evaluated in a scientific and evidence-based manner, it is important that these evaluations are performed either by independent researchers who do not have a stake in the outcome, or—if factors exist that could be perceived as conflict of interest—any such potential conflicts of interests are fully disclosed and properly managed from the outset (such as data vetting by an independent party). In the case of these two papers, the initial disclosure was incomplete, and it is unclear to what degree the competing interests of the authors were properly managed when the research was conducted.

In the JMIR Research Protocols [1], and subsequent results paper [2], authors describe a comparative analysis comparing a group of people associated with a specific “complementary medicine health care organization” (Universal Medicine, UM), with the general population, concluding in their results paper that the UM group has “unusual health indicators” (more favorable than the general population).

Both submitted manuscripts originally contained conflict of interest (COI) statements which read as follows:

> CS and VM are insiders in that they attend Universal Medicine events. However, they have received no funding, reimbursement, or other consideration from UM or its stakeholders, and no instructions or directions of any kind from UM or its stakeholders. No other competing interests exist.

After acceptance, our freelance copyeditor edited this statement out and replaced it with our standard verbiage “Conflicts of Interest: None Declared,” which is used when there is no COI, because “attending events” in itself is not normally something that would be considered a conflict of interest requiring disclosure. Authors did not object to these copyediting changes and approved the galley proofs. Their signed “license to publish” does not contain any further COI disclosures.

Shortly after publication, we received a 12-page letter from a third party whistleblower, detailing extensive undisclosed conflicts of interests of the authors, which made clear that their COIs go way beyond being “insiders in attending Universal Medicine events.” The letter was also addressed to another journal which published another protocol of the group [3], as well as to the University of Queensland (the lead author CS is associated with that university in his capacity as a PhD student).

We asked authors to provide a more detailed conflict of interest statement for a possible correction of the original papers.

In response, the lead author submitted a 1-page revised COI statement (see below) detailing that all four authors have varying degrees of association with Universal Medicine and are—most significantly—members of the “Practitioners’ Association” which is the body regulating practitioners who are “qualified to practice Universal Medicine modalities.” Of special significance is that two authors have “occasionally offered paid private healing sessions.”

The revised COI by the author also claims that “all authors have experienced substantial health benefits since they started visiting UM events.” In addition, they all have published blogs on UM associated websites. The wife of the lead author is—according to the revised COI—involved in “voluntary activities around producing content” for a UM-associated company and is a “company secretary” of the UM-associated company Unimed Living (owned through another company by the UM-founder Serge Benhayon) and “does this in an honorary capacity. She is not a director or shareholder” and “does not receive any financial incentives” from UM.

We consulted the original peer-reviewers of the results paper showing them the updated COI and they stated they would not have accepted the manuscript would they have known about these extensive COIs.

We suggested to the authors that we feel that given the significant COIs (as well as the statistical errors in the results paper, which inflated the effect sizes) both articles should be retracted and we would prefer to do this with their consent. The lead author rejected this with the argument that they originally submitted a conflict of interest which the journal removed. We maintain that the original COI submitted stating that two authors “attend UM events” was inadequate and unclear, and did not cover the full extent of the COI. The lead author CS also maintains that the involvement of his wife as company secretary for a Universal Medicine company is irrelevant because it is not a paid position. We checked the company registration documents of Unimed Living and CS’s spouse is indeed listed as company secretary, which is considered an “officer” of a corporation in Australia, so this is not just a merely administrative position, rather, they have many of the same duties and obligations as directors [4]. Thus—even in the absence of remuneration—such involvement constitutes a significant COI.

We remind our authors of the fact that “The potential for conflict of interest can exist whether or not an individual believes that the relationship affects his or her scientific judgment.” [5] and that—while financial relationships are the easiest to identify—conflicts can occur for other reasons, such as religious beliefs, personal relationships, and intellectual passion.

Our concerns with the COI of the lead author (and his spouse) go beyond financial COIs, as in his blog the lead author describes how meeting the UM founder “changed our lives profoundly” [6], and his spouse is describing “seemingly miraculous changes” [7] as a result of UM. This level of “passion” for UM and their involvement may affect the authors’ scientific judgement.

The University of Queensland has launched an investigation, but the investigation is (as of May 11th, 2018) not complete. In the meantime, we are publishing the updated COI statement as corrigendum and this statement of editorial concern, while we await the outcome of the university investigation to decide on further steps.

We are furthermore concerned about the fact that the authors recently also requested the removal of the University of Birmingham as affiliation of one co-author (JK), which is a unusual request.

While authors never claim otherwise, we should stress that the proposed [1] and executed research [2] does not provide any evidence that any Universal Medicine modalities are effective in making people healthier. There are severe limitations regarding on what can be concluded from an observational, cross-sectional study without a control group. One possible explanation for why UM members are apparently healthier than the rest of the population is simply selection bias, meaning that people being associated with UM were always healthier, or less healthy when they joined UM, with "regression to the mean" over time. Another possible explanation involves confounding factors, or the simple fact that UM members adopt healthier lifestyles.

G. Eysenbach

Editorial Director, JMIR Publications

Authors’ Corrigendum

(as submitted by the authors)

Affiliation

The authors request to change Jane Keep's affiliation to:

> The Leaders Leader, Greater London, United Kingdom

Instead of:

> Health Services Management Centre, School of Social Policy, University of Birmingham, Birmingham, United Kingdom

Conflicts of Interest

The authors were advised to change the conflict of interest statement. The new Conflict of Interest Statement should read as follows:

> All four authors have varying degrees of association with Universal Medicine and are currently members of the Esoteric Practitioners’ Association (EPA) which is the body regulating practitioners who are qualified to practice Universal Medicine modalities.
>
> Universal Medicine has a focus on complementary-to-medicine practices, that aim to support and augment medical treatments.
>
> Jane Keep has attended Universal Medicine workshops since October 2003. Jane Keep was a director of Universal Medicine UK until 2013. She is a member of the EPA, and a committee member of the EPA, and has been accredited by the EPA to offer Esoteric Healing Modalities since 2010. From 2009-2012 Jane ran a small clinic in England which offered Universal Medicine healing modalities. Since 2012 Jane has been working in corporates/universities/hospitals and occasionally offered paid private Esoteric Healing sessions, though since 2014 she has offered no paid private Esoteric Healing sessions. She was a contributor to Unimed Living 2013 – 2016. Jane has a PhD which referenced the work of over 300 people including Serge Benhayon.
>
> Eunice Minford is a Consultant General Surgeon, and has trained as an Interfaith Minister and Spiritual Counsellor. She also attended the National University of Ireland and obtained a degree of “Master of Applied Christian Spirituality” studying Sacred Esoteric Healing in her thesis. Eunice is also editor of the website “Medicine and Serge Benhayon” and a contributor to that website and to the “Unimed Living” website. She has her own blog “The Soulful Doctor” where she discusses, et al, Universal Medicine. She is also on the EPA professional committee as well as a medical advisor to, and the International Patron of, the EPA. She is a trained esoteric healing practitioner and provides occasional private sessions.

> Christoph Schnelle is a financial adviser and has some Universal Medicine associated persons among his client base. Christoph is currently working towards his PhD with The University of Queensland, the subject of which is two randomised controlled trials of Esoteric Connective Tissue Therapy (a Universal Medicine modality) on chronic low back pain and has accumulated case studies as part of this project. Christoph Schnelle's wife, Nicola Lessing, is involved in voluntary activities around producing content for "Unimed Living" and other websites. Nicola is company secretary of Unimed Living and does this in an honorary capacity. She is not a director or shareholder of Unimed Living. She is not employed by Universal Medicine or Unimed Living and does not receive any financial incentives from Universal Medicine or Unimed Living.
>
> Vanessa McHardy is involved in voluntary activities around producing content for "Unimed Living", presenting at a conference on Psychological Well Being in 2013 on the Gold Coast of Australia. She has no other involvement other than what is set out below.
>
> All four authors have experienced substantial health benefits since they started visiting Universal Medicine events. They all have published blogs on Universal Medicine associated websites and all four have commented on other blogs published on those websites.
>
> All four have no financial ties and have received no money from Universal Medicine or its related entities including no reimbursements of expenses. Each one attends more than 10 Universal Medicine events a year and regularly receive treatments from Universal Medicine accredited practitioners.

Data

In some places the standard error was used instead of the correct standard deviation.

The abstract should read:

> Differences and corresponding effect size estimates (Cohen *d*; ≥0.8 is a high difference, ≥0.5 a medium and ≥0.2 a small one with *P*<.001 except where indicated) included body mass index (BMI; 1.11), stress level (0.20, *P*=.006), depression (0.44), summary physical (0.31) and mental health (0.37), general mental health (0.39), emotional (0.15, *P*=.009) and social functioning (0.22), vitality (0.58), and general health (0.49), as well as lower incidences of diabetes, hypertension, and thrombosis (*P*<.001 each).

Instead of:

> Differences and corresponding effect size estimates (Cohen *d*; ≥0.8 is a high difference) included body mass index (BMI; 10.8), stress level (0.64), depression (4.4), summary physical (4.6) and mental health (5.1), general mental health (7.6), emotional (4.5) and social functioning (4.9), vitality (11.9), and general health (10.1), as well as lower incidences of diabetes, hypertension, and thrombosis (*P*<.001 each).

From the Data Analysis:

> We calculated Cohen *d* using the Stata command esizei (version 14.2; StataCorp LLC) using the pooled standard deviation [30] with Welch's approximation.

Should be changed from:

> We calculated Cohen *d* using the Stata command esizei (version 14.2; StataCorp LLC). The ALSWH standard deviations are reported to only 1 significant digit, so we assumed the maximum possible standard deviation (eg, 0.1 became 0.149 and 0.0 became 0.049, reducing Cohen *d*), from which we used the pooled standard deviation [30] with Welch's approximation.

From Results:

> Observed effect sizes (Cohen *d*) ranged from 0.20 to 1.11 (Table 3) [33,34].

Should be changed from:

> Observed effect sizes (Cohen *d*) ranged from 0.6 to 11.9 (Table 3). These values are with one exception higher than the 0.8 considered to denote a large effect [33,34].

The new version of Table 3 with amended Cohen *d* and *r* is as below:

Table 3. Results from standard survey scales in the Australian Longitudinal Study on Women's Health (ALSWH) and Universal Medicine (UM) groups, with *r* values and standard deviation.

	Survey respondents with ages covered by ALSWH surveys				ALSWH respondents with UM frequency weights			Effect size			
	n[a]	Mean	95% CI	SD[b]	Mean	95% CI	SD[b]	Cohen	95% CI	*r* [c]	*pc*
Body mass index (kg/m^2)	253	21.0	20.7-21.4	2.97	26.1	25.9-26.2	4.6	1.11	0.98-1.23	.48	$4*10^{-66}$
Stress[d] (lower is better)	200	0.63	0.55-0.70	0.52	0.73	0.72-0.75	0.53	0.20	0.057-0.38	.10	0.0059
Perceived Control Scale[d]	135	4.9	4.8-5.0	0.63	4.3	4.3-4.3	0.79	0.74	0.57-0.91	.35	$2*10^{-17}$
CES-D[d] (lower is better)	233	3.6	3.1-4.2	5.60	6.1	6.1-6.2	5.6	0.44	0.31-0.57	.21	$6*10^{-11}$
SF-36[d]											
Summary Physical Health	272	52.8	51.9-53.6	10.0	49.7	49.4-49.9	10.0	0.31	0.19-0.43	.15	$6*10^{-7}$
Summary Mental Health	272	51.4	50.4-52.5	10.0	47.7	47.5-47.9	10.0	0.37	0.25-0.50	.18	10^{-9}
General Mental Health	295	80.1	78.5-81.7	13.6	73.2	72.9-73.4	17.9	0.39	0.27-0.51	.19	$5*10^{-11}$
Role Emotional	294	85.3	82.2-88.3	26.5	79.6	79.2-79.9	36.9	0.15	0.038-0.27	.08	0.0091
Social Functioning	295	87.1	84.9-89.3	19.1	81.9	81.7-82.1	24.0	0.22	0.10-0.33	.11	0.0002
Vitality	295	69.5	67.6-71.5	17.2	57.5	57.2-57.8	20.7	0.58	0.47-0.70	.28	$9*10^{-23}$
General Health	275	81.9	80.0-83.8	15.9	71.8	71.6-71.9	20.9	0.49	0.36-0.61	.24	$3*10^{-15}$
Bodily Pain	294	82.8	80.6-85.0	19.5	70.7	70.4-70.9	24.0	0.51	0.39-0.62	.25	$2*10^{-17}$
Role Physical	294	84.8	81.6-88.0	27.9	78.2	77.9-78.6	36.2	0.18	0.10-0.34	.09	0.0019
Physical Function	294	89.5	87.9-91.0	13.3	84.6	84.1-85.1	19.7	0.25	0.16-0.40	.12	0.00003

[a] Number of UM respondents with ages that were surveyed in ALSWH for this particular question.

[b] SD: standard deviation.

[c] The r value was calculated as $r=d/(\text{sqrt}[4+d^2])$, where *d* is Cohen *d* as derived from the formula given by Nakagawa and Cuthill [32]. *P* value calculated with Satterthwaite's *t* test.

[d] Multi-item summed scores for perceived stress, Perceived Control Scale, Center for Epidemiologic Studies Depression Scale (CES-D), and 36-Item Short Form Survey (SF-36) using Australian coefficients.

References

1. Schnelle C, Minford EJ, McHardy V, Keep J. A Group of 500 Women Whose Health May Depart Notably From the Norm: Protocol for a Cross-Sectional Survey. JMIR Res Protoc 2017 Nov 23;6(11):e234 [FREE Full text] [doi: 10.2196/resprot.7993] [Medline: 29170148]
2. Schnelle C, Minford E, McHardy V, Keep J. Comparative Analysis of Women With Notable Subjective Health Indicators Compared With Participants in the Australian Longitudinal Study on Women's Health: Cross-Sectional Survey. JMIR Public Health Surveill 2018 Jan 10;4(1):e6 [FREE Full text] [doi: 10.2196/publichealth.9490]
3. Schnelle C, Messerschmidt S, Minford EJ, Greenaway-Twist K, Szramka M, Masiorski M, et al. Esoteric Connective Tissue Therapy for chronic low back pain to reduce pain, and improve functionality and general well-being compared with physiotherapy: study protocol for a randomised controlled trial. Trials 2017 Jul 17;18(1):328 [FREE Full text] [doi: 10.1186/s13063-017-2055-8] [Medline: 28716108]
4. Australian Institute of Company Directors. Role of the company secretary - Duties of Directors URL: http://aicd.companydirectors.com.au/resources/all-sectors/roles-duties-and-responsibilities/role-of-the-company-secretary
5. International Committee of Medical Journal Editors. Recommendations for the Conduct, Reporting, Editing and Publication of Scholarly Work in Medical Journals URL: http://www.icmje.org/ [accessed 2018-05-11] [WebCite Cache ID 6zL9p4jZl]
6. Universal Medicine Exposed. Christoph Schnelle – A Man of Truth URL: https://www.universalmedicineexposed.com/christoph-schnelle.html [accessed 2018-05-11] [WebCite Cache ID 6zLAMwY6h]
7. Universal Medicine Exposed. Nicola Lessing - Expressing Soul-fully URL: https://www.universalmedicineexposed.com/nicola-lessing.html [accessed 2018-05-11] [WebCite Cache ID 6zLAFJ2Ce]

Edited by G Eysenbach; submitted 24.01.18; this is a non–peer-reviewed article;accepted 04.04.18; published 11.05.18.

Please cite as:
JMIR Editorial Office
Expression of Editorial Concern, Correction of Conflict of Interest and Affiliation, and Data Corrections
JMIR Public Health Surveill 2018;4(2):e53
URL: http://publichealth.jmir.org/2018/2/e53/
doi:10.2196/publichealth.9932
PMID:29757750

Letter to the Editor

Overcoming Barriers to HIV Prevention and Healthcare Among Sub-Saharan African Migrants in Spain

Miriam Navarro[1,2], MD, PhD; Bárbara Navaza[1,2], MA; Anne Guionnet[1,2], MA; Rogelio López-Vélez[1,2], MD, PhD

[1]Infectious Diseases Department, National Referral Centre for Tropical Diseases, Hospital Universitario Ramón y Cajal, Madrid, Spain
[2]Salud Entre Culturas, Madrid, Spain

Corresponding Author:
Miriam Navarro, MD, PhD
Infectious Diseases Department
National Referral Centre for Tropical Diseases
Hospital Universitario Ramón y Cajal
Ctra. Colmenar, km 9,100
Madrid, 28034
Spain
Phone: 34 913368108
Email: miriamtropical@gmail.com

(JMIR Public Health Surveill 2018;4(2):e10478) doi:10.2196/10478

KEYWORDS

HIV-prevention; migrants; Africa; access to healthcare; health education

Fakoya and colleagues explored factors associated with access to HIV testing and primary care among migrants living in nine European countries—including Spain—in their study recently published in this journal [1]. The authors highlighted the importance of continued HIV knowledge and awareness initiatives aimed at migrant communities. We would like to emphasize that linguistic and cultural adaptation of such initiatives is crucial to send effective preventative messages and to overcome barriers to healthcare access and medical follow-up, especially among sub-Saharan African migrants (SSAM). Furthermore, we highly recommend the participation of intercultural mediators and we consider that the institutional support is vital to ensure the strategies' continuity.

We would like to comment on some methods and key results of the HIV prevention program carried out with migrants by the National Referral Centre for Tropical Diseases of the Hospital Ramón y Cajal in Madrid. With the aim of overcoming barriers to healthcare and HIV-prevention among migrants, the program was created in 2006 by a team of physicians, translators, intercultural mediators and a psychologist, focusing on SSAM living in Spain. From 2007, the program ("New citizens, new patients") started to cover more topics (such as Chagas disease [2], tuberculosis or travel-related diseases [3]) and to reach migrants from different continents and regions (Latin-America and the Caribbean, Eastern Europe, Maghreb, Asia).

This HIV-prevention program—still ongoing—is based on the following pillars:

1. Study and consideration of migrants' needs and perspectives (qualitative research and KAP—knowledge, attitudes and practices—questionnaires)
2. Design of educational material
3. Community engagement through conduction of:
   a. Multilingual seminars in NGOs and migrant associations
   b. Public awareness campaigns, street activities and distribution of brochures and condoms on key dates (ie, World AIDS day) and sites (ie, primary healthcare centers)
4. Overcoming barriers to HIV diagnosis and treatment by performing rapid tests and guaranteeing access to medical follow-up
5. Training of intercultural mediators
6. Training of healthcare providers and NGO professionals on HIV prevention with migrants [4]

SSAM participating in our HIV-prevention program have reported to be native-speakers of more than 30 different African languages, Wolof and Bambara being especially prominent. Given this great language diversity, we developed a training program on intercultural mediation for SSAM in 2008 (n=11; eight men and three women). Seven migrant women from other geographical regions were also trained. A second edition of the course was performed in 2014, in which seven of the 19 trainees were SSAM, all men. Currently, they act as interpreters and peer educators in informative seminars performed in several Spanish regions and in medical consultations in Madrid, playing a key role to avoid misunderstandings, overcome linguistic barriers and improve communication among patients and health professionals. Our interpretation and intercultural mediation service obtained official recognition by the Regional Government of Madrid in December 2017.

While coordinating the counseling and intercultural mediation of a pilot program offering rapid HIV testing in seven public primary healthcare centers in Madrid (2010-2013), we reached a significant number of people who had not been tested previously. Uptake of the service by SSAM was remarkable: 7.4% of migrants who used the service (51/687) [5].

Drawing from qualitative and quantitative research, our informative HIV-prevention seminars include key aspects about the functioning of the Spanish health system, as well as specific information for SSAM about blood testing. Previous experiences of SSAM with healthcare systems—both in Africa and Spain—influenced the perception of blood testing as a potentially unhealthy and abusive practice [6]. Reluctance to undergo blood testing was hampering patients' medical follow-up while increasing barriers for HIV testing and treatment. Throughout our program, we have observed that rapid diagnostic tests—both finger-prick and oral fluid tests—are usually well accepted by SSAM, hence facilitating access to HIV testing. Nevertheless, the abovementioned perceptions of routine blood testing can hamper confirmatory tests and medical follow-up of HIV patients, as we have experienced during medical consultations. For these reasons, we inform about what is done with the blood, and we explain why "so much" blood is sometimes needed, how our body recovers afterwards, how long it takes to get the results and the patients' rights to be informed about such results. Specific brochures for SSAM were developed, where the abovementioned information about blood testing was included. This material was translated into French, English and Portuguese and is freely available on the Internet, on the websiteof the NGO *Salud Entre Culturas*, working in the Hospital Ramón y Cajal [7]. Our materials are designed by healthcare professionals working along with intercultural mediators and are fully illustrated to reach people regardless of their literacy. In 2017 we performed 150 HIV rapid tests after our seminars with SSAM and three of them (2%) shown a reactive result.

Last, we would like to highlight the importance of measuring the efficacy of the interventions, in order to further adapt them to the changing target population's profile. Fakoya and colleagues reported additional individual-level obstacles to HIV prevention and testing such as lack of knowledge about HIV and low perception of risk [1]. The existence of HIV was questioned or denied by a significant number of migrants who participated in our program: 13% (487/3759; 2007-2017), being most of them SSAM. Globally, the proportion of participants who believed that HIV existed increased significantly after attending our seminars (47% vs. 95%; n=473; 2008-2011). They also reported a better understanding of HIV and AIDS (n=691; P<.001), a decrease on their own discriminatory attitudes towards HIV positive people and an increase in preventative practices such as condom use. Furthermore, the percentage of participants undergoing HIV testing rose significantly after our intervention (average time between KAPs: four weeks; n=473; P<.001).

While analyzing the factors influencing the level of knowledge about HIV using multilingual KAP questionnaires, we observed no correlation between the variable "having received previous information about HIV in Spain" and a higher level of knowledge (2006-2009) [8]. This highlighted the need of better adapting the strategies to reach migrants who did not speak Spanish or were not familiarized with the Spanish healthcare system. Thus, we started to train healthcare providers and NGO professionals on HIV prevention with migrants from 2008 [4].

Working with an interdisciplinary—and highly motivated—team which included migrants acting as intercultural mediators and peer educators was paramount for the effectiveness and acceptability of the HIV prevention program. Another key element for the positive impact of our program was the collaboration with physicians, researchers and experts in infectious diseases working in the Hospital Ramón y Cajal. We would like to conclude by encouraging public institutions to fund interdisciplinary and community-based interventions in order to assure their feasibility and continuous adaptation to the specific needs of the population.

References

1. Fakoya I, Álvarez-Del AD, Copas AJ, Teixeira B, Block K, Gennotte AF, et al. Factors Associated With Access to HIV Testing and Primary Care Among Migrants Living in Europe: Cross-Sectional Survey. JMIR Public Health Surveill 2017 Nov 06;3(4):e84 [FREE Full text] [doi: 10.2196/publichealth.7741] [Medline: 29109072]
2. Navaza B, Guionnet A, Navarro M, Estévez L, Pérez-Molina JA, López-Vélez R. Reluctance to do blood testing limits HIV diagnosis and appropriate health care of sub-Saharan African migrants living in Spain. AIDS Behav 2012 Jan;16(1):30-35. [doi: 10.1007/s10461-010-9871-0] [Medline: 21259044]
3. Navarro M, Navaza B, Guionnet A, López-Vélez R. A multidisciplinary approach to engage VFR migrants in Madrid, Spain. Travel Med Infect Dis 2012 May;10(3):152-156. [doi: 10.1016/j.tmaid.2012.03.001] [Medline: 22592114]
4. Manual práctico para realizar talleres de prevención del VIH con personas inmigrantes. URL: http://www.saludentreculturas.es/wp-content/uploads/2015/10/Manual-pr%C3%A1ctico-para-realizar-talleres-de-prevenci%C3%B3n-del-VIH-con-personas-inmigrantes.pdf [accessed 2018-04-20] [WebCite Cache ID 6ypGN0Ii7]

5. Esteban-Vasallo MD, Morán-Arribas M, García-Riolobos C, Domínguez-Berjón MF, Rico-Bermejo J, Collado-González S, et al. Targeted rapid HIV testing in public primary care services in Madrid. Are we reaching the vulnerable populations? Int J Infect Dis 2014 Feb;19:39-45 [FREE Full text] [doi: 10.1016/j.ijid.2013.10.006] [Medline: 24269650]
6. Navaza B, Guionnet A, Navarro M, Estévez L, Pérez-Molina JA, López-Vélez R. Reluctance to do blood testing limits HIV diagnosis and appropriate health care of sub-Saharan African migrants living in Spain. AIDS Behav 2012 Jan;16(1):30-35. [doi: 10.1007/s10461-010-9871-0] [Medline: 21259044]
7. Salud Entre Culturas. Triptico VIH/sida adaptado para población subsahariana URL: http://www.saludentreculturas.es/wp-content/uploads/2012/10/espa%C3%B1ol.pdf [accessed 2018-05-14] [WebCite Cache ID 6zPgPHKzo]
8. Estévez L, Navarro M, Guionnet A, Navaza B, Herrero JM, Díaz M. Adapting prevention programs: variables that influence HIV/AIDS awareness among immigrants in Spain. Oral Abstracts Tropical Medicine & International Health 2009:14-97. [doi: 10.1111/j.1365-3156.2009.02353.x]

Edited by T Sanchez; submitted 22.03.18; peer-reviewed by J Llenas, J Reyes-Urueña; comments to author 13.04.18; revised version received 03.05.18; accepted 07.05.18; published 15.05.18.

Please cite as:
Navarro M, Navaza B, Guionnet A, López-Vélez R
Overcoming Barriers to HIV Prevention and Healthcare Among Sub-Saharan African Migrants in Spain
JMIR Public Health Surveill 2018;4(2):e10478
URL: http://publichealth.jmir.org/2018/2/e10478/
doi:10.2196/10478
PMID:29764802

Viewpoint

Strengthening Routine Data Systems to Track the HIV Epidemic and Guide the Response in Sub-Saharan Africa

Brian Rice[1], BSc, MSc, PhD; Andrew Boulle[2], MBChB, PhD, MSc; Stefan Baral[3], MD, MPH, MBA; Matthias Egger[4], MSc, FFPH, DTM&H; Paul Mee[1], MSc, PhD; Elizabeth Fearon[1], MSc, PhD; Georges Reniers[5], MSc, PhD; Jim Todd[5], MSc; Sandra Schwarcz[6], MPH, MD; Sharon Weir[7], MPH, PhD; George Rutherford[6], MD, AM; James Hargreaves[1], MSc, PhD

[1]Social and Environmental Health Research, Faculty of Public Health and Policy, London School of Hygiene and Tropical Medicine, London, United Kingdom

[2]Centre for Infectious Disease Epidemiology and Research, Faculty of Health Sciences, University of Cape Town, Cape Town, South Africa

[3]Department of Epidemiology, Johns Hopkins School of Public Health, Baltimore, MD, United States

[4]Institute of Social and Preventive Medicine, University of Bern, Bern, Switzerland

[5]Department of Population Health, Faculty of Epidemiology and Population Health, London School of Hygiene and Tropical Medicine, London, United Kingdom

[6]Global Health Sciences, University of California San Francisco, San Francisco, CA, United States

[7]Department of Epidemiology, University of North Carolina, Chapel Hill, NC, United States

Corresponding Author:
Brian Rice, BSc, MSc, PhD
Social and Environmental Health Research
Faculty of Public Health and Policy
London School of Hygiene and Tropical Medicine
15-17 Tavistock Place
London, WC1H 9SH
United Kingdom
Phone: 44 (0)2079272567
Email: brian.rice@lshtm.ac.uk

Abstract

The global HIV response has entered a new phase with the recommendation of treating all persons living with HIV with antiretroviral therapy, and with the goals of reducing new infections and AIDS-related deaths to fewer than 500,000 by 2020. This new phase has intensive data requirements that will need to utilize routine data collected through service delivery platforms to monitor progress toward these goals. With a focus on sub-Saharan African, we present the following priorities to improve the demand, supply, and use of routine HIV data: (1) strengthening patient-level HIV data systems that support continuity of clinical care and document sentinel events; (2) leveraging data from HIV testing programs; (3) using targeting data collection in communities and among clients; and (4) building capacity and promoting a culture of HIV data quality assessment and use. When fully leveraged, routine data can efficiently provide timely information at a local level to inform action, as well as provide information at scale with wide geographic coverage to strengthen estimation efforts.

(JMIR Public Health Surveill 2018;4(2):e36) doi:10.2196/publichealth.9344

KEYWORDS

HIV; data; systems; surveillance; sub-Saharan Africa; testing; treatment; prevention; cascade; monitoring; quality assessment; clinical; program

Background

The global HIV response has entered a new phase with the recommendation of treating all persons living with HIV with antiretroviral therapy [1], and with the goals of reducing the number of new HIV infections and AIDS-related deaths to fewer than 500,000 by 2020 [2]. This new phase has intensive data requirements to monitor progress toward these goals.

The infrastructure of national-level public health surveillance platforms across the most heavily impacted countries, particularly in sub-Saharan Africa, is improving. This presents an opportunity for a major shift in the type of data used to inform country-level HIV prevention, treatment, and care responses. Central to this shift are recommendations by the World Health Organization (WHO) for investing in strategies to collect and leverage routine HIV data to track the HIV epidemic and ultimately, the response [3]. Moreover, new laboratory technologies, the growth of networked information systems, and the development of sophisticated data collection and analysis methods offer the potential to develop better strategic information.

Concurrent with these advances are shifts away from the mainstays of HIV. It has been suggested that the ethical underpinnings of unlinked anonymous HIV testing of pregnant women in sentinel surveillance have weakened, particularly in areas where HIV status information may be extracted from routine medical records [4-6]. National household-level surveys require large investments and provide only periodic data on which to anchor trend estimates. There remains a widespread absence of HIV incidence data at a time when the interpretation of prevalence trends is becoming ever more complex as treatment continues to expand and survival continues to lengthen. Intensive community-based cohort studies provide critical insights on epidemic dynamics but only over a limited geographic scope [7]. Sentinel HIV treatment cohorts and clinical databases are compromised by challenges in ascertaining outcomes and ensuring follow-up [8].

To promote sustainability, operational and surveillance platforms that utilize the latest computer technology for data storage are required. These platforms will ideally provide information in a timely fashion and at a scale and degree of detail that will improve our understanding of the distributions of HIV risk and determinants of that risk across settings. Similarly, we need surveillance systems that can monitor progress towards reducing the undiagnosed fraction of persons living with HIV, accelerate their pathway to care, and support retention in care. A greater focus on strengthening the demand, supply, analysis, and use of routine HIV data is essential to meet these needs and guide prevention, testing, and treatment efforts.

Focusing on sub-Saharan Africa, we present the following priorities for action to improve routine HIV data: (1) strengthening individualized HIV data systems that support continuity of clinical care and document sentinel events; (2) leveraging data from HIV testing programs; (3) promoting targeted routine data collection in communities and among clients; and (4) building capacity and promoting a culture of HIV data quality assessment and use (Multimedia Appendix 1). We define routine data as data generated, through regular procedure, from service-delivery platforms for prevention, testing and treatment, community and client surveys designed to inform to inform programmatic or strategic needs, and national registries. These data may be paper-based or electronic (preferably electronic) and presented aggregately or individually (preferably individually). All of the surveillance data recommended in this paper can and should be collected routinely.

Priorities for Action

Strengthen Individualized HIV Data Systems

Although the HIV care continuum [9-12] and 90-90-90 fast-track strategy [13] have given energy to measuring progress and driving scale-up in treatment programs, there remains much to do to ensure progress is accurately tracked. In many sub-Saharan African countries, patient monitoring data come from multiple sources, are of varying quality, and are often recorded electronically long after they have been combined to create aggregate reports. As these systems frequently monitor a person's care pathway only from commencement of treatment, the pathway from HIV infection to diagnosis and to treatment is poorly understood, as is the wider clinical experience. Most current district health information systems rely on aggregate data making it difficult to assess how complete or accurate the data are or whether duplicate records are present [14].

Track Patients Within and Across Programs

Central to maximizing the utility of routine data is the idea of patient tracking. Data systems that track individuals over time and space are the only practicable way to ensure good clinical care and program responses. The ability to link individuals in this way enables the identification of persons lost to follow-up from one facility who have silently transferred to another facility, or who have died in the community, thereby informing targeted patient tracing or back-to-care programs [15]. These same systems also offer invaluable opportunities for strategic information based on de-duplicated individual-level records to compliment traditional reporting through management information systems in which the data are aggregated at source, and improve national and global estimates.

Individual patient data leveraged from multiple systems can provide an important data source for developing a comprehensive strategic HIV information system, such as case surveillance [3,14]. To expand existing HIV surveillance activities, the WHO has recommended adopting or strengthening such systems [3]. When collated through a comprehensive strategic HIV information system these patient data can provide a count of HIV infected and diagnosed individuals across a specified geographic area, and provide a comprehensive and timely picture of the care cascade. A comprehensive strategic HIV information system is but one component among a suite of components to strengthen the use of patient data in surveillance. Wherever possible, extracting data from systems already in place for operational and clinical purposes is preferred to requiring clinicians to additionally report events for surveillance purposes.

Develop Unique Universal Personal Identifiers

To facilitate system linkage, better methods are required for the development and integration of unique, universal, health identifiers that work across multiple health services and systems, such as linkage to vital registration, tuberculosis and viral hepatitis control programs, and maternal, new-born, and child health systems [3,8,16,17]. Lessons learned from countries with substantial experience with unique identifiers should be considered to inform adoption or expansion in other countries.

Approaches may include civil identification numbers, national health identifiers or master patient indexes (a database used by multiple locations to consistently maintain information on each registered patient), or developing bespoke matching algorithms to link records based on existing patient identifying information (eg, sex, age, and a representation of name such as a Soundex) [18]. In developing unique identifiers, it is essential to consult with and address the concerns of clients and civil society on the potential for confidentiality breaches and consequent human rights abuses.

Utilize Data From HIV Testing Platforms

The growing array of mechanisms for delivering HIV testing in clinical and non-clinical settings offers great potential for epidemiological intelligence. However, realizing this potential requires careful thought and investment.

Re-Focus on Data Systems for New Diagnoses

Data collected through HIV testing platforms, both negative and positive results, can inform prevention and clinical service provision. Site-specific information on newly diagnosed persons, including sex, age (as derived from date of birth), CD4 cell count and, where feasible, probable route of infection and a marker of residence, can be used to identify geographical regions and groups where HIV incidence is elevated. Spatial maps can be produced from these data to inform the provision of targeted and to-scale prevention and treatment services. Information on persons testing HIV negative can be utilized to inform appropriate prevention service provision. Developing data quality assurance protocols can help ensure the HIV testing data we collect are comprehensive and robust.

Testing data are critical for monitoring early (CD4 cell count greater than 350 cells/µL) and, conversely, late diagnosis (less than 350 cells/µL). As low CD4 cell counts can be triggers for more intensive follow-up [19], testing data can be used to monitor differentiated care models and promote earlier diagnosis. Testing data are also critical for monitoring the pathway from diagnosis to care, an important and often ignored step in the care cascade. In addition, to objectively monitor performance across the 90-90-90 targets, a first stage denominator of overall HIV prevalence is required. To monitor the first and second 90s we require an estimate of persons living with HIV who have been diagnosed for the numerator and denominator, respectively. If systems are strengthened, routine testing data should in the future directly provide estimates of diagnosed HIV, and indirectly through modeling, of undiagnosed HIV in sub-Saharan African settings.

Incorporate Data from Multiple Testing Platforms

Data collected through facility, community, and home-based testing platforms, key population outreach programs and, increasingly, through the expansion of self-testing all offer the potential for informing action. To utilize these data, a better understanding of the impact of shifting combinations, and increased coverage, of testing provision on reducing the number of previously untested persons and the number undiagnosed is needed, as is gaining a better understanding of the drivers for testing outside of traditional clinical settings.

The ongoing shift for surveillance purposes from unlinked anonymous testing in antenatal sentinel sites to using routine antenatal testing data collected through existing HIV program monitoring systems provides an opportunity to improve the level of detail by which HIV prevalence estimates are presented [20,21]. If coverage of HIV testing services at antenatal clinics is high, and if the available data are of high quality, this shift can improve the representativeness of HIV prevalence estimates as they are among all pregnant women rather than just among women from a non-representative sample of health facilities. However, there is a need to conduct quality assurance assessments of these data as there are remaining concerns that routine data, as compared to sentinel site data, may present greater variation in coverage and quality [22-24].

The expanding number of mechanisms for routinely delivering HIV tests ensures methodological improvements are required to appropriately analyze and interpret the data generated at the local and national level. It is likely that, amongst others, care-seeking patterns, existing knowledge of HIV status, test-kit quality, missing patient responses, and a variable prevalence of electronic data systems and standardized practices will influence estimates.

Employ Recency Assays

Understanding where and among whom new infections are occurring is critical to plan programs and to track the incidence of new infections over time to assess whether control measures are working. To date, most estimates of HIV incidence in sub-Saharan Africa come from modeling efforts based on trends in HIV prevalence, births and deaths, and information about the coverage of treatment programs. In some settings estimates that do not rely on modeling are based on community cohort studies with repeat HIV testing, although this has rarely, with some exceptions [25], been done at national scale.

Recent infection testing algorithms (RITA) can distinguish recently acquired infection from long-standing infection [26,27]. Recency testing offers the potential to inform action for individual patients, their networks and programs and, if incorporated into routine testing and programmatic surveys, to strengthen estimates of incidence. Examples might include identifying clusters of infection among key populations in outreach testing platforms, or complimenting data on HIV prevalence in routine testing of pregnant women with information on recency. Attention to test accuracy, data quality, ethics, sample size, and analysis and interpretation will be necessary. Nevertheless, the wider adoption of these technologies within routine data systems, especially if the price can be reduced, has the potential to bring new strategic information.

Targeted Routine Data Collection Among Communities and Clients

Strategically deployed community-based and client surveys (eg, tracing studies) can routinely be conducted to complement our understanding gleaned through prevention, testing, and treatment programs. They can inform and correct some of the biases of routine programmatic data, and provide information about individuals and populations at risk of infection or treatment failure, the scale and nature of their needs, and potential barriers to the effectiveness of prevention programs [28]. These surveys are often cost efficient compared to population-wide surveys and provide an opportunity to contextualize populations represented within routinely collected service data, and to develop HIV prevention cascades to guide prevention programs [29,30]. Because these populations are often hidden and stigmatized, and may practice, in some cases, illegal behaviors, novel methods are required.

Systematize Key Population Surveillance

In sub-Saharan Africa, venue-based and respondent-driven sampling strategies are commonly used to characterize key populations for whom there is no practical sampling frame. Venue-based sampling focuses on settings where key populations congregate, while respondent-driven sampling leverages networks of people sharing risk characteristics [31-33]. Although widely used in Africa, these approaches have not to date received the same level of attention as national-level general population surveys.

These sampling methods tend to be based in local sites and often have not been designed to be representative at larger scale. Current practice has led to a range of funders and programs deploying a variety of sampling methods. Heterogeneity in methods and sample populations limit the potential for the data to be used to track trends or be incorporated in modeling efforts designed to inform resource allocation decisions.

The implications of which sampling approach is used has been shown to be significant, with the characteristics of the populations recruited differing [33]. Through gaining a greater understanding of the strengths, limitations, and interpretation of sampling methods, a more systematic approach to their deployment can be developed. Decision-trees can be adopted to highlight which assumptions were met, which were not, and how best to proceed given different circumstances and statistical tools developed to derive best estimates from data arising from routine services and targeted surveys.

Identify Populations With Greatest Needs and Develop Tools to Guide Service Provision

Surveillance for program improvement requires identifying the location, needs and size of populations at high risk, and recognizing individuals may move in and out of periods of high risk and/or fall under one or more high-risk behavioral group. It is essential that program decision-making is based not only on those reached by services but also those unreached.

There is a growing call for community-based efforts that differentiate sub-groups and risk to inform targeted service provision. Within the Determined, Resilient, Empowered, AIDS-Free, Mentored, and Safe Women (DREAMS) initiative, the Girl Roster tool is being deployed to help reach adolescent girls in need, while variants of network-based referral can also be used to identify young women who sell or trade sex for referral into programs; approaches such as peer-led micro-planning can be deployed to strengthen programming [34-36].

Build Capacity and Promote a Culture of HIV Data Quality Assessment and Use

As good local data spur action locally and create better data for onward reporting, the systems designed ought to be sustainable and responsive to the local context and produce data that are user-friendly, sustainable, and meet the needs of stakeholders at various levels. Long-term investment in human and technological capacity is required so that people within highly impacted countries can create their own solutions and develop sustainable practice. Collectively, we should avoid demanding more data unless this can be derived without impact on data collectors or front-line clinicians.

Encourage Responsibility for Data quality at the Local Level

Data quality at the local-level can be improved through promoting timely, comprehensive and accurate facility-based record keeping, matching data needs to setting, and building buy-in and trust among stakeholders. Engagement with and training of frontline data staff are critical. Understanding the needs of clinical staff is also critical. If data systems are strengthened for clinical care governance, data quality and strategic information will benefit. Deriving population-representative strategic data from routine clinical and/or health systems will also reduce the potential for data collection for surveillance purposes becoming a separate resource-intensive activity in itself.

Evaluate Data Quality

At the sub-national and national level, a standardized approach to system design, reducing redundancy, and limiting the development of parallel systems (systems with a shared purpose developed independent of one another) should be encouraged. Through reviewing existing systems and/or surveillance frameworks [37,38], elements of transferable practice can be identified to inform system design and the development of training materials.

The engagement and training of data stewards at the sub-national and national-level (eg, district and national Minister of Health staff) to oversee programmatic, clinical, and surveillance activities are critical. These custodians can promote robust feedback loops so that the common problem of data travelling only in one direction, from facility to health departments in aggregate form, can be avoided. Feedback to those who collect, collate, and report data should rank reporting sites according to data quality and program outcomes. Feedback should incorporate the use of maps and dashboards, and make clear what the benefits of using the data provided have been (eg, informing policy and interventions).

To ensure data outputs are interpreted correctly, assessments of data sources are required to indicate quality and value. For example, to compare results across populations and time for each stage of the care cascade, a traffic light system could be used to indicate confidence levels in data sources of high, medium, and low. To generate high quality data, and promote accurate interpretation, conducting quality assessments and analyses to investigate and correct some of the biases of routine data is also needed. Bias may arise, from missing data, non-representativeness, double counting (aggregate data particularly suffer from this), or loss to follow-up. Through identifying bias, methods such as adjustment factors, to support the interpretation of routine data outputs, can be designed.

Develop Procedural Documents

At all levels, standard operational procedures for data collection and quality assurance should be developed, promoted, and updated regularly. These procedural documents can specify how routine data quality assessments are to be conducted and instruct how identified errors are addressed. They can stipulate standardized analytical methods for producing useful programmatic, clinical and surveillance outputs that account for potential biases and include accurate assessments of error, uncertainty, and explicit consideration of important assumptions. These documents can also provide guidance for conducting regular output evaluations to ensure accuracy, to ensure the needs of users are being met, and to monitor impact.

Consider Timeliness of HIV Data Cycles

Data cycles need to be feasible and designed to meet the different needs of a variety of stakeholders. For clinical activities, it is required that data be collected, reviewed, and shared as close to real time as possible. Such provision of data can enable missed appointments to be acted upon immediately. For surveillance activities, data producers will ideally collate as few variables as necessary, and only those that inform action. They will also assess the minimum period required for conducting comprehensive data collection and quality improvement, and producing and reviewing meaningful outputs. In conjunction with this, data users should assess the timeframes within which they can receive, review, and use the data for program improvement. The development and endorsement of standard operating procedures can guide these efforts.

Discussion

To reduce HIV incidence and mortality in sub-Saharan Africa we need to provide effective prevention methods, identify all individuals living with undiagnosed HIV infection, improve the pathway between diagnosis and care, retain people once they are in care, and promote antiretroviral therapy adherence. To achieve this a new phase must be entered, one where we innovate and improve the routine data systems that support HIV prevention and care and where we shift from using routine data purely for descriptive purposes towards directly informing and improving clinical care and program performance. Robust data to inform advocacy, to secure resources, and to create an environment conducive to better rights and fit within national monitoring and evaluation policies are required.

Here, we presented four priorities that will result in durable and replicable data systems that collect, store, and produce individual-level data, and that give rise to user-friendly tools that can be used by local, national, and international practitioners to drive more effective and efficient clinical management and prevention and care programming. We focus on routine data that are, or might be, collected in sub-Saharan Africa as part of service delivery requirements. We are united in our appreciation of the value of national surveys of HIV prevalence and incidence, community-based cohort studies, and research studies and trials, but our focus is now on routine HIV data systems that will be central to the next wave of HIV strategic information. While we focused on HIV-specific data, we also recognize the enormous value of strengthening vital registration and broader health information systems.

In promoting the use of routine data we need to assess the level of disaggregation useful to inform action and recognize that uncertainty and/or imprecision increases with the level of disaggregation. We also need to collect and present data by geographies that map to how programs and services are delivered.

When fully leveraged, routine data can efficiently provide timely information at a local level to inform action, as well as provide information at scale with wide geographic coverage to strengthen estimation efforts. To maximize the use of these data we need to overcome complex challenges. We need to promote sustainability through investment in data system design, implementation and capacity, and to ensure the data we produce are credible and valuable through strong attention to data quality and use, and through careful analysis and investigation of potential biases.

Acknowledgments

We gratefully acknowledge the expert group of public health practitioners who gathered in Tallinn, Estonia in October 2016 for the Measurement and Surveillance of HIV Epidemics (MeSH) Consortium Scientific Symposium on Data Driven Decision Making to Control the HIV Epidemic. The key points, solutions, and recommendations presented in this paper reflect the author's consideration of the discussions held at the Tallinn symposium. This work is funded by the MeSH Consortium.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Summary of priorities for action.

[PDF File (Adobe PDF File), 311KB - publichealth_v4i2e36_app1.pdf]

References

1. World Health Organization. Consolidated guidelines on the use of antiretroviral drugs for treating and preventing HIV infection: recommendations for a public health approach - second edition. Geneva: World Health Organization; 2016. URL: http://www.who.int/hiv/pub/arv/arv-2016/en/ [accessed 2018-03-06] [WebCite Cache ID 6xiOCa4Ru]
2. UNAIDS. 2016 United Nations Political Declaration on Ending AIDS sets world on the fast-track to end the epidemic by 2030. Geneva: UNAIDS; 2016. URL: http://www.unaids.org/en/resources/presscentre/pressreleaseandstatementarchive/2016/june/20160608_PS_HLM_PoliticalDeclaration [accessed 2018-03-15] [WebCite Cache ID 6xvxi82yW]
3. World Health Organization. Consolidated guidelines on person-centred HIV patient monitoring and case surveillance Geneva: World Health Organization. Geneva: World Health Organization; 2017. URL: http://apps.who.int/iris/bitstream/10665/255702/1/9789241512633-eng.pdf [accessed 2018-03-15] [WebCite Cache ID 6xvxpyNJv]
4. Dee J, Garcia CJ, Marsh K, Zaidi I, Murrill C, Swaminathan M. HIV surveillance among pregnant women attending antenatal clinics: evolution and current direction. JMIR Public Health Surveill 2017 Dec 05;3(4):e85 [FREE Full text] [doi: 10.2196/publichealth.8000] [Medline: 29208587]
5. Rennie S, Turner AN, Mupenda B, Behets F. Conducting unlinked anonymous HIV surveillance in developing countries: ethical, epidemiological, and public health concerns. PLoS Med 2009 Jan 20;6(1):e4 [FREE Full text] [doi: 10.1371/journal.pmed.1000004] [Medline: 19166264]
6. Gonese E, Mushavi A, Mungati M, Mhangara M, Dzangare J, Mugurungi O, et al. Is Zimbabwe ready to transition from anonymous unlinked sero-surveillance to using prevention of mother to child transmission of HIV (PMTCT) program data for HIV surveillance?: results of PMTCT utility study, 2012. BMC Infect Dis 2016 Feb 29;16:97 [FREE Full text] [doi: 10.1186/s12879-016-1425-2] [Medline: 26923185]
7. Reniers G, Wamukoya M, Urassa M, Nyaguara A, Nakiyingi-Miiro J, Lutalo T, et al. Data resource profile: Network for Analysing Longitudinal Population-Based HIV/AIDS Data on Africa (ALPHA Network). Int J Epidemiol 2016 Feb;45(1):83-93 [FREE Full text] [doi: 10.1093/ije/dyv343] [Medline: 26968480]
8. Petersen M, Yiannoutsos C, Justice A, Egger M. Observational research on NCDs in HIV-positive populations: conceptual and methodological considerations. J Acquir Immune Defic Syndr 2014 Sep 01;67 Suppl 1:S8-16 [FREE Full text] [doi: 10.1097/QAI.0000000000000253] [Medline: 25117964]
9. Gardner EM, McLees MP, Steiner JF, Del RC, Burman WJ. The spectrum of engagement in HIV care and its relevance to test-and-treat strategies for prevention of HIV infection. Clin Infect Dis 2011 Mar 15;52(6):793-800 [FREE Full text] [doi: 10.1093/cid/ciq243] [Medline: 21367734]
10. Hull M, Wu Z, Montaner J. Optimizing the engagement of care cascade: a critical step to maximize the impact of HIV treatment as prevention. Curr Opin HIV AIDS 2012 Nov;7(6):579-586. [doi: 10.1097/COH.0b013e3283590617] [Medline: 23076123]
11. Alemnji G, Fonjungo P, Van Der Pol B, Peter T, Kantor R, Nkengasong J. The centrality of laboratory services in the HIV treatment and prevention cascade: the need for effective linkages and referrals in resource-limited settings. AIDS Patient Care STDS 2014 May;28(5):268-273 [FREE Full text] [doi: 10.1089/apc.2013.0356] [Medline: 24742299]
12. McNairy ML, El-Sadr WM. The HIV care continuum: no partial credit given. AIDS 2012 Sep 10;26(14):1735-1738. [doi: 10.1097/QAD.0b013e328355d67b] [Medline: 22614888]
13. UNAIDS. 90-90-90: an ambitious treatment target to help end the AIDS epidemic. Geneva: UNAIDS; 2014. URL: http://www.unaids.org/sites/default/files/media_asset/90-90-90_en_0.pdf [accessed 2018-03-15] [WebCite Cache ID 6xvy9hRfo]
14. Harklerode R, Schwarcz S, Hargreaves J, Boulle A, Todd J, Xueref S, et al. Feasibility of establishing HIV case-based surveillance to measure progress along the health sector cascade: situational assessments in Tanzania, South Africa, and Kenya. JMIR Public Health Surveill 2017 Jul 10;3(3):e44 [FREE Full text] [doi: 10.2196/publichealth.7610] [Medline: 28694240]
15. Zürcher K, Mooser A, Anderegg N, Tymejczyk O, Couvillon M, Nash D, IeDEAMESH consortia. Outcomes of HIV-positive patients lost to follow-up in African treatment programmes. Trop Med Int Health 2017 Dec;22(4):375-387 [FREE Full text] [doi: 10.1111/tmi.12843] [Medline: 28102610]
16. Boulle A, Van Cutsem G, Hilderbrand K, Cragg C, Abrahams M, Mathee S, et al. Seven-year experience of a primary care antiretroviral treatment programme in Khayelitsha, South Africa. AIDS 2010 Feb 20;24(4):563-572. [doi: 10.1097/QAD.0b013e328333bfb7] [Medline: 20057311]
17. Fox M, Brennan A, Maskew M, MacPhail P, Sanne I. Using vital registration data to update mortality among patients lost to follow-up from ART programmes: evidence from the Themba Lethu Clinic, South Africa. Trop Med Int Health 2010 Apr;15(4):405-413 [FREE Full text] [doi: 10.1111/j.1365-3156.2010.02473.x] [Medline: 20180931]
18. Mortimer J, Salathiel J. 'Soundex' codes of surnames provide confidentiality and accuracy in a national HIV database. Commun Dis Rep CDR Rev 1995 Nov 10;5(12):R183-R186. [Medline: 8541940]
19. World Health Organization. Guidelines for managing advanced HIV disease and rapid initiation of antiretroviral therapy. Geneva: World Health Organization; 2017. URL: http://www.who.int/hiv/pub/guidelines/advanced-HIV-disease/en/ [accessed 2018-03-15] [WebCite Cache ID 6xvyFyJyR]

20. World Health Organization. Guidelines for conducting HIV surveillance among pregnant women attending antenatal clinics based on routine programme data. Geneva: World Health Organization; 2015. URL: http://www.unaids.org/en/resources/documents/2015/SurveillanceRoutineProgrammeData [accessed 2018-03-15] [WebCite Cache ID 6xvyOCXMq]
21. World Health Organization. Consolidated guidelines on HIV testing services: 5Cs: consent, confidentiality, counselling, correct results and connection. Geneva: World Health Organization; 2015. URL: http://apps.who.int/iris/handle/10665/179870 [accessed 2018-03-15] [WebCite Cache ID 6xvyW7MkQ]
22. Boeras D, Peeling R. External quality assurance for HIV point-of-care testing in Africa: a collaborative country-partner approach to strengthen diagnostic services. Afr J Lab Med 2016;5(2):556 [FREE Full text] [doi: 10.4102/ajlm.v5i2.556] [Medline: 28879132]
23. World Health Organization. Improving the quality of HIV-related point-of-care testing: ensuring the reliability and accuracy of test results. Geneva: World Health Organization; 2015. URL: http://www.who.int/hiv/pub/toolkits/handbook-point-of-care-testing/en/ [accessed 2018-03-15] [WebCite Cache ID 6xvyb95y6]
24. Sheng B, Marsh K, Slavkovic AB, Gregson S, Eaton JW, Bao L. Statistical models for incorporating data from routine HIV testing of pregnant women at antenatal clinics into HIV/AIDS epidemic estimates. AIDS 2017 Apr;31 Suppl 1:S87-S94. [doi: 10.1097/QAD.0000000000001428] [Medline: 28296804]
25. Nsanzimana S, Remera E, Kanters S, Mulindabigwi A, Suthar A, Uwizihiwe J, et al. Household survey of HIV incidence in Rwanda: a national observational cohort study. Lancet HIV 2017 Oct;4(10):e457-e464. [doi: 10.1016/S2352-3018(17)30124-8] [Medline: 28801191]
26. Mastro TD, Kim AA, Hallett T, Rehle T, Welte A, Laeyendecker O, et al. Estimating HIV incidence in populations using tests for recent infection: issues, challenges and the way forward. J HIV AIDS Surveill Epidemiol 2010 Jan 01;2(1):1-14 [FREE Full text] [Medline: 21743821]
27. Aghaizu A, Murphy G, Tosswill J, DeAngelis D, Charlett A, Gill ON, et al. Recent infection testing algorithm (RITA) applied to new HIV diagnoses in England, Wales and Northern Ireland, 2009 to 2011. Euro Surveill 2014 Jan 16;19(2) [FREE Full text] [Medline: 24457006]
28. World Health Organization. Consolidated guidelines on HIV prevention, diagnosis, treatment and care for key populations. Geneva: World Health Organization; 2014. URL: http://www.who.int/hiv/pub/guidelines/keypopulations/en/ [accessed 2018-03-15] [WebCite Cache ID 6xvyjTEEI]
29. Hargreaves JR, Delany-Moretlwe S, Hallett TB, Johnson S, Kapiga S, Bhattacharjee P, et al. The HIV prevention cascade: integrating theories of epidemiological, behavioural, and social science into programme design and monitoring. Lancet HIV 2016 Dec;3(7):e318-e322. [doi: 10.1016/S2352-3018(16)30063-7] [Medline: 27365206]
30. Garnett G, Hallett T, Takaruza A, Hargreaves J, Rhead R, Warren M, et al. Providing a conceptual framework for HIV prevention cascades and assessing feasibility of empirical measurement with data from east Zimbabwe: a case study. Lancet HIV 2016 Dec;3(7):e297-e306 [FREE Full text] [doi: 10.1016/S2352-3018(16)30039-X] [Medline: 27365204]
31. Kendall C, Kerr L, Gondim R, Werneck G, Macena R, Pontes M, et al. An empirical comparison of respondent-driven sampling, time location sampling, and snowball sampling for behavioral surveillance in men who have sex with men, Fortaleza, Brazil. AIDS Behav 2008 Jul;12(4 Suppl):S97-104. [doi: 10.1007/s10461-008-9390-4] [Medline: 18389357]
32. Weir S, Merli M, Li J, Gandhi A, Neely W, Edwards J, et al. A comparison of respondent-driven and venue-based sampling of female sex workers in Liuzhou, China. Sex Transm Infect 2012 Dec;88 Suppl 2:i95-101 [FREE Full text] [doi: 10.1136/sextrans-2012-050638] [Medline: 23172350]
33. Rao A, Stahlman S, Hargreaves J, Weir S, Edwards J, Rice B, et al. Sampling key populations for HIV surveillance: results from eight cross-sectional studies using respondent-driven sampling and venue-based snowball sampling. JMIR Public Health Surveill 2017 Dec 20;3(4):e72 [FREE Full text] [doi: 10.2196/publichealth.8116] [Medline: 29054832]
34. Closing the gender gap - how efforts to collect data about women and girls drive global economic and social progress. New York: New York Times; 2017. URL: https://paidpost.nytimes.com/gates-foundation/closing-the-gender-data-gap.html [accessed 2018-03-15] [WebCite Cache ID 6xvyqYHdK]
35. Avahan - the India AIDS initiative: the Business of HIV prevention at scale. New Delhi: Bill & Melinda Gates Foundation; 2008. URL: https://docs.gatesfoundation.org/documents/avahan_hivprevention.pdf [accessed 2018-03-15] [WebCite Cache ID 6xvz2jmZf]
36. Cowan FM, Davey CB, Fearon E, Mushati P, Dirawo J, Cambiano V, et al. The HIV care cascade among female sex workers in Zimbabwe: results of a population-based survey from the Sisters Antiretroviral Therapy Programme for Prevention of HIV, an Integrated Response (SAPPH-IRe) Trial. J Acquir Immune Defic Syndr 2017 Dec 01;74(4):375-382. [doi: 10.1097/QAI.0000000000001255] [Medline: 27930599]
37. van Dam ASG, Valenzuela BHB. The surveillance system of the centers sexual health - Netherlands. Netherlands: Netherlands National Institute for Public Health and the Environment; 2017. URL: https://rivm.nl/Documenten_en_publicaties/Algemeen_Actueel/Uitgaven/Infectieziekten_Bulletin/Jaargang_28_2017/Maart_2017/Inhoud_maart_2017/Het_surveillancesysteem_van_de_Centra_Seksuele_Gezondheid [WebCite Cache ID 6yC0fw8jA]
38. HIV surveillance systems. London: Public Health England; 2008. URL: https://www.gov.uk/guidance/hiv-surveillance-systems [accessed 2018-03-15] [WebCite Cache ID 6xvzNFb2T]

Abbreviations

WHO: World Health Organization
MeSH: Measurement and Surveillance of HIV Epidemics

Edited by T Sanchez; submitted 06.11.17; peer-reviewed by J Dee, C Murrill; comments to author 23.11.17; revised version received 08.02.18; accepted 14.02.18; published 03.04.18.

Please cite as:
Rice B, Boulle A, Baral S, Egger M, Mee P, Fearon E, Reniers G, Todd J, Schwarcz S, Weir S, Rutherford G, Hargreaves J
Strengthening Routine Data Systems to Track the HIV Epidemic and Guide the Response in Sub-Saharan Africa
JMIR Public Health Surveill 2018;4(2):e36
URL: http://publichealth.jmir.org/2018/2/e36/
doi:10.2196/publichealth.9344
PMID:29615387

Viewpoint

Sustainable Monitoring and Surveillance Systems to Improve HIV Programs: Review

Daniel Low-Beer[1]; Mary Mahy[2], ScD; Francoise Renaud[1], PhD; Txema Calleja[1]

[1]World Health Organization, Geneva, Switzerland
[2]UNAIDS, Geneva, Switzerland

Corresponding Author:
Daniel Low-Beer
World Health Organization
20 avenue Appia
Geneva, 1211
Switzerland
Phone: 41 795090693
Email: daniellowbeer@gmail.com

Abstract

HIV programs have provided a major impetus for investments in surveillance data, with 5-10% of HIV program budgets recommended to support data. However there are questions concerning the sustainability of these investments. The Sustainable Development Goals have consolidated health into one goal and communicable diseases into one target (Target 3.3). Sustainable Development Goals now introduce targets focused specifically on data (Targets 17.18 and 17.19). Data are seen as one of the three systemic issues (in Goal 17) for implementing Sustainable Development Goals, alongside policies and partnerships. This paper reviews the surveillance priorities in the context of the Sustainable Development Goals and highlights the shift from periodic measurement towards sustainable disaggregated, real-time, case, and patient data, which are used routinely to improve programs. Finally, the key directions in developing person-centered monitoring systems are assessed with country examples. The directions contribute to the Sustainable Development Goal focus on people-centered development applied to data.

(JMIR Public Health Surveill 2018;4(2):e3) doi:10.2196/publichealth.8173

KEYWORDS

surveillance; HIV; development monitoring; evaluation

Introduction

The United Nations (UN) Sustainable Development Goals (SDGs) launched in 2015 are a universal set of 17 goals, targets, and indicators that United Nations member states are expected to use to frame their agendas and political policies. These SGDs have consolidated HIV, tuberculosis (TB), malaria, and other communicable diseases into one target, Target 3.3, and health into one goal, Goal 3 [1,2].

At the same time, data have increasing prominence in the SDGs and are seen as one of the three systemic issues in implementation (Goal 17) [3]. There are two data targets, for (1) reliable, timely, disaggregated data (Target 17.18) and (2) capacity for data use, namely statistical and analytic capacity for countries to measure their own progress (17.19).

Data were largely highlighted for reporting against targets in the UN's Millennium Development Goals (MDGs) of 2000. In the SDGs, routine data use has a much more active, ongoing role as part of implementation—the so-called data revolution. The shift in the role of data from the MDGs to the SDGs is from reporting after implementation, to using data upfront as part of how development is implemented. As the Independent Expert Advisory Group report on Mobilising the Data Revolution for sustainable development states, "Data are the lifeblood of decision-making and the raw material for accountability" [4].

In fact, HIV programs have made major contributions to investments in data; for example, the US President's Emergency Plan for AIDS Relief (PEPFAR) program has supported population-based surveys with, more recently, surveys including incidence measurement. The Global Fund to fight acquired immune deficiency syndrome (AIDS), TB, and malaria recommends that 5-10% of its US $4 billion funds each year is invested in monitoring and evaluation (M&E) and has specific strategic initiatives to invest in district health information systems and key population data [5]. However, the SDG priority to use data routinely leads to three major directions for the surveillance agenda.

First, countries need to develop sustainable, routine patient and case surveillance systems based on program records since measures of prevalence from national population-based surveys are prohibitively expensive for most countries.

Second, although there are 100 indicators defined by the World Health Organization (WHO) for health, there are only five cross-cutting data systems, which generate the majority of these data [6]. There is increasing agreement among partners to align support to the basic data systems, where data can then be used to improve programs. Alignment of partners to support data systems may be more effective than the MDG focus on agreeing indicators.

Third, there is a need for more disaggregated, individual level data, which can be used for ongoing program improvement. Routine data systems can be used for both patient care and the majority of surveillance needs [7-9].

This paper reviews an important shift towards routine monitoring and surveillance systems in the context of the wider M&E of the HIV response. This paper reviews the global surveillance priorities agreed by HIV partners in 2015 at the onset of the SDGs [10]. Second, the increasing use of sustainable, routine data systems for health care and surveillance needs is highlighted. Finally, the key directions to develop sustainable surveillance systems over the next 5 years, including person-centered case surveillance and patient monitoring systems, are discussed [11].

A Common HIV Surveillance Agenda: Towards Sustainable Data Systems

In 2015, the WHO and the Joint United Nations Programme on HIV/AIDS (UNAIDS) held the 3rd Global Consultation Meeting on HIV surveillance aiming to review country surveillance needs, assess priority gaps, and consolidate a global surveillance agenda from 2015 to 2020 [10]. This country and partner meeting coincided with the shift from the MDGs to the SDG agenda.

The WHO country-led platform for information and accountability [12] highlighted that underlying the many indicators, there were five key data collection systems (Figure 1). These were identified as population-based surveys, facility assessments, administrative sources, clinical reporting systems, and civil registration. The data systems provided a more consolidated focus than indicators for partner alignment in monitoring and surveillance.

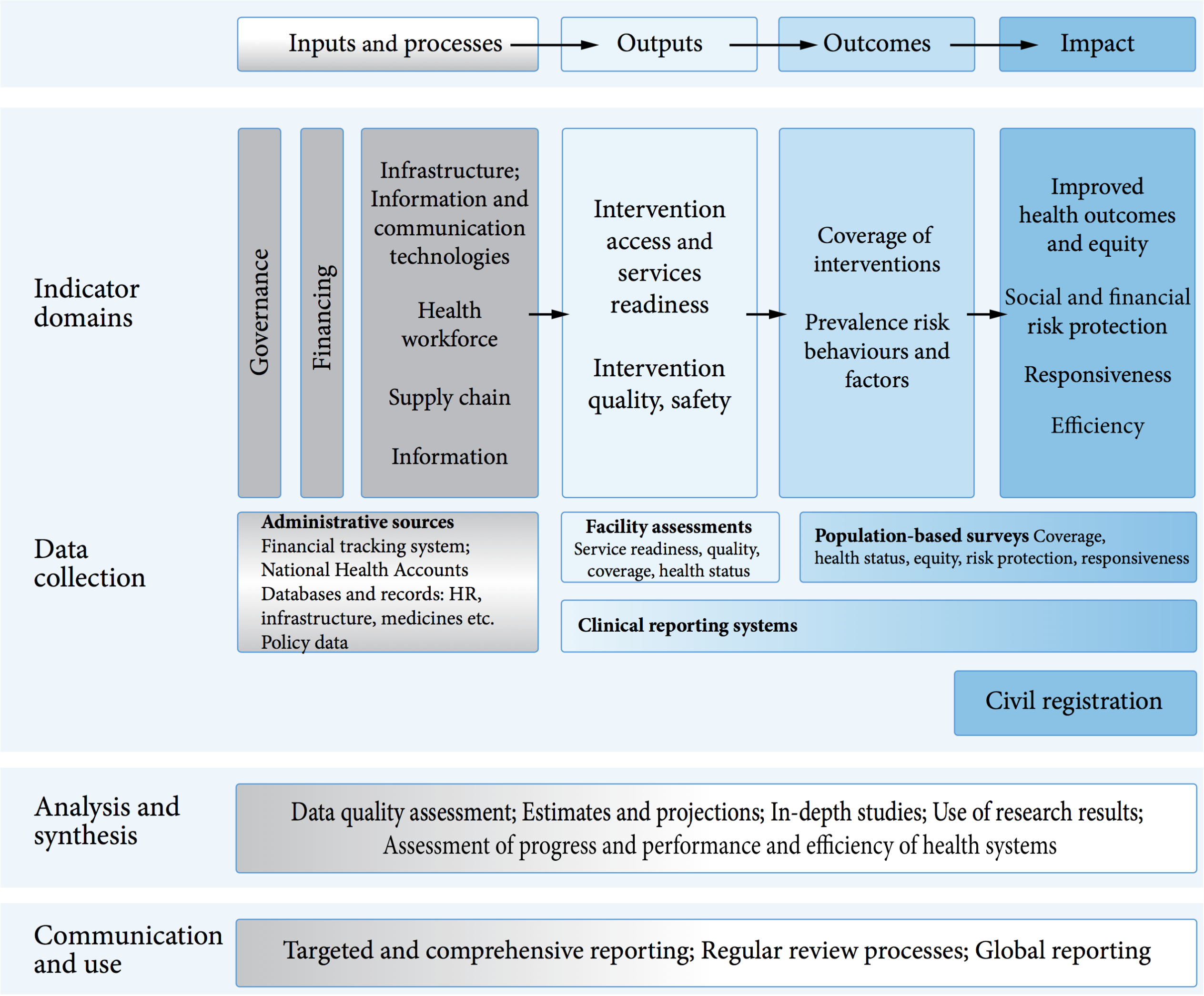
A critical balance is required between the periodic use of population-based surveys and the more routine systems for measuring facility, clinical, and civil registration (and administrative) events [13]. Initially, AIDS case surveillance had ensured data on diagnosis were collected in facilities as part of health workers’ ongoing patient care, supporting direct engagement with people living with HIV and care from their families and communities. This routine data played an important role in early community and prevention responses [14,15]. Increasingly, HIV surveillance data have been taken out of the health settings where the diagnosis and care of individuals occurs.

In 2015, it was important to rebalance support of all five components of data collection in the context of the SDGs. The global HIV surveillance agenda was developed by partners to support a shift towards more sustainable surveillance. It also highlighted some key gaps and challenges, including for key populations, to directly measure HIV incidence, to standardize and improve the quality of routine individual-level data, and to use routine data more actively for program improvement.

The country and global partners agreed on the role of surveillance data and a shift in balance towards routine, integrated surveillance data at a local level, based on the following principles:

1. A key component of ending AIDS epidemic is better quality, local, granular, and disaggregated data to design and support a sustainable response.
2. Supporting the health services cascade requires a cascade of linked data.
3. The use of surveillance data is an intervention in itself allowing programs and communities to better respond to the epidemic with services.
4. Surveillance requires systematic investments of at least 5-10% of program funds so that overall funds are focused on the epidemic and so that impact can be assessed.
5. Increased support is needed for routine, integrated, district health data as part of health information systems including sexually transmitted infections and hepatitis, linked to real-time health decisions.

Figure 1. Common M&E framework defined by WHO.



After reviewing the evidence of developments in surveillance methods and country implementation, five key surveillance priorities were agreed on at the Global Surveillance Meeting in 2015, set for the period 2015-2020 [10]):

1. HIV prevalence (and incidence) data that are granular and disaggregated to the local level, by age, sex, and populations
2. Key population data that inform program and national estimates, including stigma, behaviors, and linkage to services
3. HIV case reporting and facility data that monitor coverage and linkages from prevention to treatment and to other health areas
4. Measurement and review of mortality and incidence, including modeling and estimation approaches, and program impact reviews
5. Analysis capacity to use multiple large datasets from various sources, including data from surveys, facilities and communities, and from new media, for advocacy, program improvement, and estimating impact

Surveillance priorities for 2015-2020 were agreed on that described the guidance required and status of implementation (Table 1). This provides an aligned partner and country approach to surveillance data in HIV, which has often been fragmented [6,8]. In addition, regular epidemiological reviews provide an input to wider country program reviews, providing a greater focus on impact.

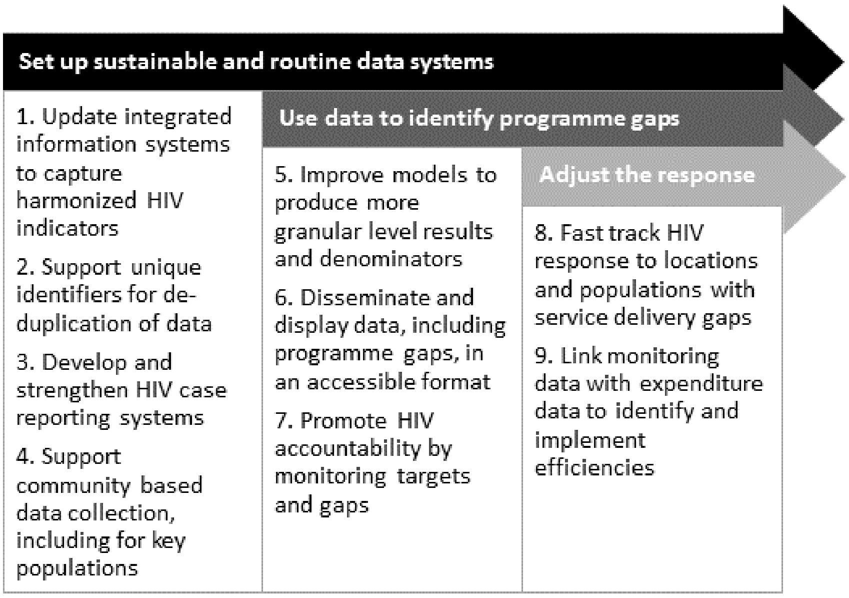
Based on the global surveillance agenda, WHO and UNAIDS have increased support for sustainable routine data systems and use [16]. To achieve the fast track targets, UNAIDS has proposed a shift in strategic information systems (Figure 2). The first four elements of Figure 2 shows the foundation for a sustainable routine monitoring system described in this paper.

The proposed shift in strategic information strengthens the focus on community-based data collection and links to finance and expenditure data. By analyzing the granular routine data with expenditure data and modeling where required, countries will have the strategic information necessary to fast-track the HIV response in the geographic areas and populations that are most in need. [17].

Table 1. Global surveillance priorities and work plan for development.

Key area	Guidance status	Gaps to fill
1. Incidence	Yes, for household surveys, gaps for routine data	Guidance on incidence assays for surveys Application for case diagnoses
2. Mortality	Yes, for civil registrations and vital statistics, and for demographic sentinel surveillance sites	Guidance on sentinel or routine data on HIV-related mortality
3. Household surveys	Guidance provided	Need for update on household surveys
4. Key population data	Yes	Bio-behavioral surveys, guidelines to complete Strategic framework for use of data on key populations, from program to national levels Guidance on size estimate algorithm, use for local and national programs
5. Case surveillance	Yes, in draft	Guidance on person-centered monitoring, patient and case surveillance, and use of unique identifiers in HIV and health
6. ANC routine testing data	Shift from antenatal clinic sentinel sites to prevention of mother to child transmission of HIV and use of routine testing data	Gap for implementation support for use of testing data for surveillance and also for new infections
7. Analysis capacity	Yes, need to support cascade gap analysis	Impact reviews and prioritization: new sources of impact data Big data: new analysis methods for facility and program data

Figure 2. Shift in strategic information to fast track the HIV response (developed by UNAIDS and supported by WHO).



Set up sustainable and routine data systems

1. Update integrated information systems to capture harmonized HIV indicators
2. Support unique identifiers for de-duplication of data
3. Develop and strengthen HIV case reporting systems
4. Support community based data collection, including for key populations

Use data to identify programme gaps

5. Improve models to produce more granular level results and denominators
6. Disseminate and display data, including programme gaps, in an accessible format
7. Promote HIV accountability by monitoring targets and gaps

Adjust the response

8. Fast track HIV response to locations and populations with service delivery gaps
9. Link monitoring data with expenditure data to identify and implement efficiencies

Developing Routine Monitoring Systems for Patient Care and Surveillance

WHO has convened partners including Centers for Disease Control and Prevention (CDC), UNAIDS, Global Fund, civil society, and countries to consolidate guidance on person-centered patient monitoring and case surveillance data. This aims to consolidate routine data systems to be used for improved care and for the majority of reporting needs.

The aim of this guidance is to combine routine, real-time patient and program monitoring and surveillance into a consolidated M&E system. This involves defining standardized sentinel events (from diagnosis, treatment to viral suppression, and outcomes, which are reported in the same way across a program) for consistent reporting. The guidelines extend this surveillance system from patients, to all cases diagnosed with HIV, and with unique identifiers to link to other health services.

The development of these guidelines was based on a country situation analysis of sustainable country monitoring systems at different levels of development, regions, and contexts (Table 2).

The guidelines define a sustainable routine monitoring system based on (1) promoting the use of routine data for patient care and enabling reporting on most program, national, and global indicators in a sustainable manner, (2) supporting linkage of HIV patient care to wider health care needs and monitoring, including mortality data, through the use of unique identifiers and the principles of interoperability of data systems, (3) building on ongoing, routine monitoring systems in countries to strengthen country capacity, and (4) providing disaggregated, person-centered data to improve services and to support the ongoing, real-time use of data at local level for improved health care and program improvement.

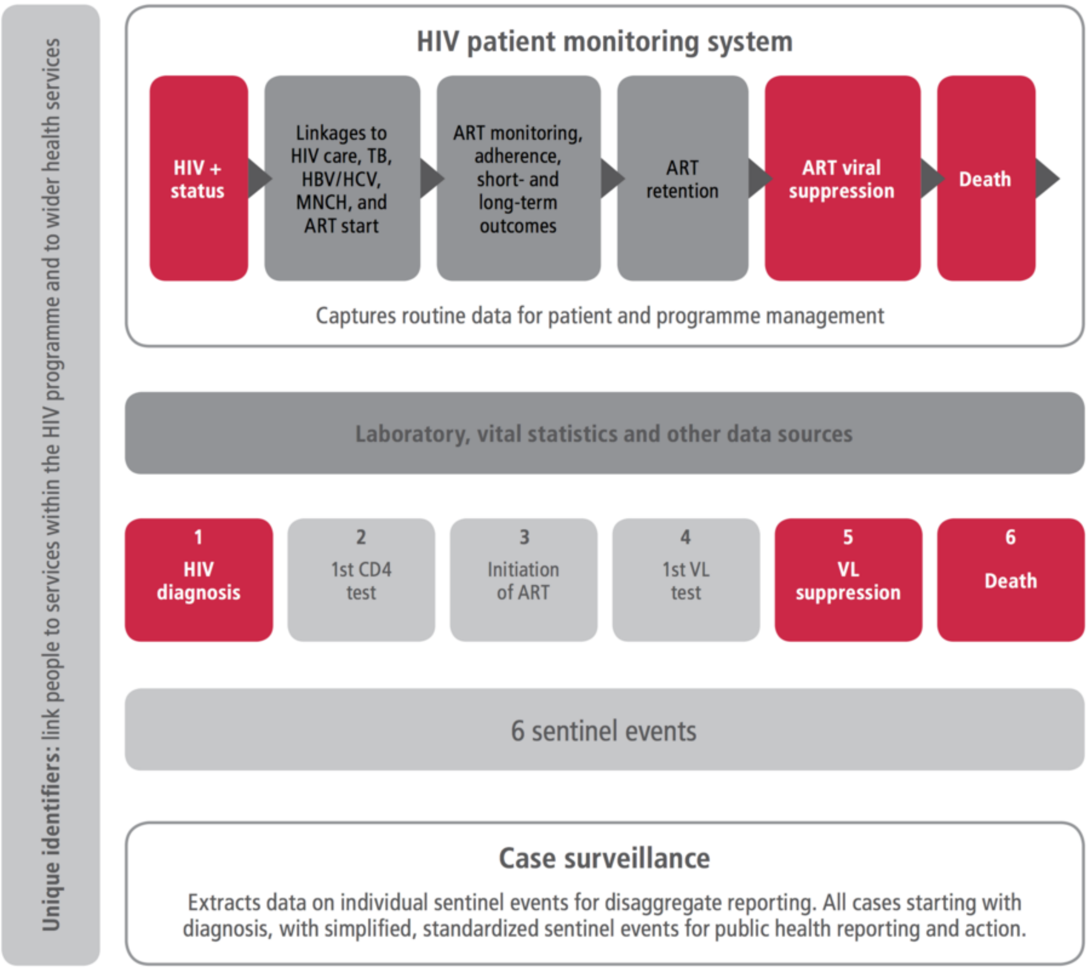
While developing patient monitoring systems improve patient health care, major risks of confidentiality and security of data also need to be assessed in each context. The guidelines aim to strengthen investments in the robustness, interoperability, and security of routine data systems. The approach produced in the recent WHO guidelines is in line with the SDGs on people-centered development and "leaving no one behind," where data are disaggregated to ensure that, for example, key populations and groups by age, gender, and geography receive services.

The approach to strengthening routine data for patient monitoring and surveillance consists of three key components, as presented in the guidelines on person-centered monitoring: (1) use of standardized processes and tools for HIV patient monitoring at the health facility level, linked to priority indicators for national and global level reporting, (2) strengthening of case-based surveillance for HIV to integrate all cases of HIV based on key sentinel events starting with diagnosis, and (3) use of unique identifiers to link patients and their related health data between health services, for example, to link people with HIV on lifetime treatment to wider chronic care services (Figure 3).

Table 2. Situation analysis of countries developing person-centered case surveillance and patient monitoring.

Country	Surveillance system	Program improvements and gaps
Haiti	Individual case surveillance introduced with single national dataset integrating multiple sources. Data de-duplicated and used to identify transfers. Minimal cost, as built on existing infrastructure and data.	Targeted HIV treatment services as populations migrated seasonally. Better directed prevention resources. Generates routine reporting.
Zimbabwe	Building case surveillance on patient monitoring system. 80% of records contain unique identification of national insurance number. Need to invest in a robust and secure macro database to link facilities.	Major benefits for retention and contacting those lost to follow-up, removing those who have gone to other facilities or who have died. Need to invest in a robust and secure macro database to link facilities.
Brazil	Primary case reporting in place built for payment purposes, not surveillance. Labs require CD4 and viral load to receive payment from Ministry of Health. Uses names and includes key population information to assess equal access.	Works well and improves follow-up and payment. Major limitation does not include private laboratories. Assess access to key populations, ensure confidentiality and human rights protection.
Zambia	Smartcard system used to link patient records but does not cover all facilities. Not all facilities linked online; data collected on memory sticks from some sites.	Major benefit of being able to de-duplicate testing and treatment records, for improved patient management and more accurate reporting.
Malawi	Health "passport" for all health services. Differentiated system in which all HIV sites with more than 2000 patients use electronic medical records, but most sites are still paper-based. Data are entered into electronic database centrally.	Quarterly reporting from routine system for management, and major benefits for drug forecasting. Next step to integrate HIV with national identification and health passport.
Thailand	Unique identification based on social insurance, links key databases for patient management.	Improved availability and speed of lab test results, improved reimbursement. Migrants not covered by national unique identification.
Botswana	Routine use of national unique identification and insurance number for access to all HIV, health, and social services	Easier access, transfer and linkage to a range of HIV and health services.
Western Cape, South Africa	Three-tiered system with paper at lowest level, entered into electronic register at district level, and electronic records in 15 sites. Tier.net in 3000 sites which feeds back to patient management.	Regular, routine reports to facilities on loss to follow-up, viral load data to improve patient care and de-duplicate data.
Myanmar	Patient reporting system initially based on non-governmental organization (NGO) programs delivered by Médecins Sans Frontières. Challenge is transition to national system with investments in patient index, interoperability and links to health information system software.	Strong data on treatment cascade routinely used to highlight gaps and improve late initiation of antiretroviral therapy (ART). Facilitates planning and global reporting.

Figure 3. A consolidated routine M&E system for HIV patient monitoring and surveillance of key program measures (yellow boxes show existing indicators in patient monitoring, red shows measures requiring considerable additional investment to improve and adjust data systems).



A consolidated, routine data system needs to link data from different sources focused on data use (Figures 1, 3 and 4). This is facilitated by the introduction of unique identifiers so different data can be consolidated to support the person receiving services over their lifetime (and provide the basis for chronic care for HIV and wider healthcare). The data can provide routine information on most indicators for HIV reporting.

However, further work is needed to strengthen the reporting on community data, key populations, and prevention services provided outside facilities to inform and guide country-led prevention and control efforts. There are innovative approaches in a number of countries to link community and facility data into a single reporting system, for example in Kenya for prevention, in Zimbabwe linking sex worker outreach data to initiation of care, in India with community prevention mapping, and in Cambodia to focus on HIV incidence reduction between prevention and treatment. However, this area requires considerable strengthening as identified by UNAIDS [12,18].

The guidelines that WHO, UNAIDS, CDC, PEPFAR, and Global Fund support provide the basis of a sustainable routine, person-centered M&E system. The guidelines provide 15 recommendations to develop this approach, based on maturing and strong investments in existing country data systems, data security, and analytical capacity in countries (Table 3).

The guidance provides the backbone for sustainable routine monitoring and surveillance systems.

Table 3. Recommendations for person-centered patient monitoring and case surveillance.

Recommendations	Supporting tools (online annexes)
Patient monitoring	
1. Collect a minimum, standardized dataset for patient care.	Guidance on a minimum dataset for patient monitoring.
2. Transition monitoring to “treat all”: Depending on national guidelines, countries should transition from using the pre-ART register to using the ART register.	Guidance for this transition.
3. Simplify and standardize tools (cards, registers, and reports) across facilities.	Generic tools for adaptation.
4. Integrate and link HIV and health reporting; the HIV card should form part of the patient folder or passport integrated with primary health.	Generic HIV patient card and ART register for country adaptation.
5 Implement regular data quality reviews and invest in data use.	Guidance on carrying out an annual patient monitoring review and improving quality of care.
Case-based surveillance	
1. Standardize reporting of sentinel events: Standardized sentinel events should be identified to include the 6 key sentinel events (HIV diagnosis, first CD4 test, initiation of ART, first viral load test, viral load suppression, mortality).	Definitions of six key sentinel events
2. De-duplicate testing and treatment data to support facilities and improve data quality: Case-based surveillance should provide de-duplicated counts of diagnosed persons and people on treatment for reporting and to be shared with facilities.	Guidance on approaches.
3. Develop case surveillance based on a country situation analysis. Improvements to case-based surveillance should be based on a country situation analysis that identifies and costs incremental improvements, and not introduced as a separate monitoring approach.	Tool for country situation analysis.
4. Start case surveillance with HIV diagnosis and build on patient monitoring.	Guidance on HIV case definitions and case surveillance; requires reporting on HIV diagnosis in addition to and linked to treatment data.
5. Ensure confidentiality and security of all data, particularly for key population data. The guidance suggests that risk behavior and key population data be assessed at the point of diagnosis and to support referral to care. However, it is not routinely included in patient monitoring, where there are risks.	Recommendations on key population data
Scaling-up unique identifiers for person-centered monitoring	
1. Introduce and use unique identifiers for data shared across a program.	Definitions and examples of unique identifiers.
2. Transition progressively from paper-based to electronic patient information systems. Countries should use a tiered approach starting with high volume sites.	Example of a tiered approach.
3. Strengthen and differentiate data security: significant investments are now required in databases and policies to protect and differentiate security and confidentiality of key data.	Guidance on key components of strengthening and differentiating data security
4. Invest in data systems and promote interoperability and open source standards.	5-10% of program budgets are used to strengthen monitoring and evaluation.
5. Use data to improve programs, to strengthen retention, linkage and transfer. Data use drives data and program improvement.	Investments in data analysis functions and dashboards, which feedback data and can convene and measure program improvements.

Figure 4. Key data sources to support reporting on the cascade of services.

Routine Community Data for Key Populations

Integrating community level data into district, province, and national surveillance and monitoring systems is fundamental for shifting to sustainable routine data systems. Community-based organizations provide care and treatment services for key populations linked to essential prevention activities. The provision of these services are an important component of the HIV response in a country and must be considered when identifying populations not reached by prevention and treatment services.

At the program level, key population surveillance systems can benefit from many of the components of routine data systems. However, if the policy and data environments are not protective, it may be safer for this data to be kept at NGO and program level for the benefits of delivering services. In this case, no individual level data from the key population services will be available in the national monitoring system.

Further work is needed to strengthen the reporting of community data on key populations and prevention services provided outside of facilities. Routine data must be supplemented by regular surveys to assess HIV prevalence, incidence, and behavior to provide insight into whether the data collected in the routine systems are representative.

Routine HIV Prevalence Data From Antenatal Clinics

Since the early 2000s, UNAIDS and WHO have recommended that countries conduct sentinel surveillance among pregnant women attending antenatal clinics to determine trends in HIV over time. However these surveys require personnel and financial resources that are not often available without outside funding. These surveys were conducted every 2 years and were usually targeted at high-volume antenatal clinics often in urban areas. These data were critical for understanding the trajectory of the HIV epidemic historically.

As more countries started testing all pregnant women attending antenatal clinics, these routinely collected data have started to be used for surveillance purposes. Since 2017, it is possible to include these routine data in the modeling software that countries use to produce estimates of HIV incidence, prevalence, and AIDS-related deaths. These data allow for more granular models and allow countries to produce estimates every year instead of every 2 years.

Priority Actions and Investments in Data

The Global Surveillance Plan, 2015-2020 (Table 1) identified the following investment priorities to strengthen surveillance systems:

1. The strengthening of incidence and mortality methods and data, which can be used or with mortality data linked to routine surveillance
2. Strengthening of key population M&E, including surveys and size estimation. Many of the benefits of routine data in key populations are even more valuable for services and program management. However, issues of confidentiality and data security can be even more critical in these populations in certain settings.
3. Investing in analysis capacity, so that data are used for program improvement. The consolidated strategic

information guidelines also called for data analysis positions or a small team in the Ministry of Health to respond to the demand for data for program improvement.

4. Support transition to using routine HIV testing data from antenatal clinics for routine prevalence surveillance. This is a key component of routine surveillance for all diagnosed cases of HIV and to develop approaches to measure HIV prevalence from this data.
5. In addition, periodic household surveys will still be important to supplement routine monitoring data, particularly in a limited number of high-burden countries. However, they should be supplemental rather than replace the use of sustainable routine surveillance data in most countries.

Conclusion

The context of the SDGs provides both impetus and new directions for the surveillance agenda. There is a greater emphasis on data as a major component of implementation, at a time when HIV funding from donors is no longer increasing.

First, there is a shift towards more routine, disaggregated, real-time data. Routine data will be supplemented by national population-based surveys with biomarkers in a limited number of high-burden countries. Strong information systems based on routine data will allow further disaggregation of data (at least by age, sex, and geographic area) in accordance with SDG target 17.18.

Second, investments will need to be made in analytical capacity to better standardize, communicate, and use these data to improve programs. The value of data is in its use to improve programs at national, subnational, facility, and community level. This requires the human capacity for analysis and use alongside the hardware of M&E systems.

Finally, the value of consolidated, sustainable routine surveillance systems is for reporting but also for ongoing, real-time improvements to programs. In the coming few years, the evidence for the benefits of data as an intervention in its own right will need to be collected. This will be key to the medium-term sustainability of these remaining routine monitoring and surveillance systems.

Patient monitoring and case surveillance provide ongoing, real-time data on the cascade of health services. However, investments will be needed in the security of data systems, their interoperability, and the capacity to use the data at national, district, community, and facility levels. This is aligned with the SDG Target 17.19 on country analytic and statistical capacity.

Third, it is important that no one is left behind with these data systems. This will require additional focus on integrating monitoring data from programs working with key populations and communities on prevention, and vital events data into a basic dataset that can be used for district level program decisions.

There are limitations to a reliance on routine data for both health care and reporting needs. Routine data are prioritized for sustainability but are also a necessity given limitations to funding. Routine surveillance should be supplemented by additional national and key population surveys, cohorts (valuable for estimates, incidence, and longitudinal cascade measures). Modeling is also critical for estimating indicators that are not easily measured, synthesising multiple source of data (routine data, surveillance data, survey data, demographic data) to better understand the epidemic and the impact of the HIV response. At the same time, surveys and modeling should not substitute for routine surveillance data, which will be the backbone for reporting needs.

Programs should also focus on routine program data because of its value for ongoing, everyday program decisions. The focus on routine data use in this paper puts surveillance data back into program settings where it can be used for millions of ongoing decisions by frontline health providers. Crucial to the incentives for quality reporting by health providers will be data feedback to all levels and use for program improvement. The sustainability of M&E systems will depend on data use in program settings for documented program improvement.

Routine surveillance data, including AIDS case surveillance, have been a critical component to mobilize the HIV care and prevention response to support health care workers, families and the community [19]. Over time, data generally have been taken out of the health care setting and community, with anonymous testing, national surveys, and a reliance on modelling. This paper suggests that routine data are part of implementation and that data are an important part of the public health intervention package of prevention, care, and data. Without routine data it is very difficult to have a public health or community response.

It is important that key surveillance data are put back into routine health care and community settings, as the SDGs regard data use as an intervention. This paper supports the use of ongoing, routine monitoring data for surveillance and program implementation. Person-centered, routine case and patient monitoring can contribute to the people-centered development focus of the SDGs, applied in a sustainable manner to data.

Conflicts of Interest

None declared.

References

1. Low-Beer D. Managing for results: a “common currency” to coordinate health development. In: Low-Beer D, editor. Innovative Health Partnerships: The Diplomacy of Diversity. Singapore: World Scientific Publishing Group; 2012:117.

2. World Health Organization. World Health Statistics 2016: Monitoring health for the SDGs. Geneva, Switzerland: WHO; 2016. URL: http://www.who.int/gho/publications/world_health_statistics/2016/en/ [accessed 2017-12-12] [WebCite Cache ID 6vep3psBZ]
3. United Nations. Transforming our world: the 2030 agenda for global action. New York, NY: United Nations; 2015. URL: https://www.un.org/development/desa/disabilities/news/news/disability-and-the-sustainable-development-goals.html [accessed 2017-12-12] [WebCite Cache ID 6vepZkLxc]
4. United Nations Secretary-General's Independent Expert Advisory Group on a Data Revolution for Sustainable Development. A World That Counts: Mobilising the Data Revolution for Sustainable Development.: Data Revolution Group; 2014. URL: http://www.undatarevolution.org/wp-content/uploads/2014/11/A-World-That-Counts.pdf [accessed 2017-12-12] [WebCite Cache ID 6veq4KLJy]
5. Low-Beer D, Beusenberg M, Hayashi C, Calleja T, Marsh K, Mamahit A, et al. Monitoring HIV Treatment and the Health Sector Cascade: From Treatment Numbers to Impact. AIDS Behav 2017 Jul;21(Suppl 1):15-22 [FREE Full text] [doi: 10.1007/s10461-017-1754-1] [Medline: 28401415]
6. Global Reference List of 100 Core Health Indicators. Geneva, Switzerland: World Health Organization; 2015. URL: http://apps.who.int/iris/bitstream/10665/173589/1/WHO_HIS_HSI_2015.3_eng.pdf?ua=1 [accessed 2017-12-12] [WebCite Cache ID 6veqSM2GP]
7. Bekker L, Venter F, Cohen K, Goemare E, Van Cutsem G, Boulle A, et al. Provision of antiretroviral therapy in South Africa: the nuts and bolts. Antivir Ther 2014;19 Suppl 3:105-116. [doi: 10.3851/IMP2905] [Medline: 25310359]
8. Govindasamy D, Ford N, Kranzer K. Risk factors, barriers and facilitators for linkage to antiretroviral therapy care: a systematic review. AIDS 2012 Oct 23;26(16):2059-2067. [doi: 10.1097/QAD.0b013e3283578b9b] [Medline: 22781227]
9. Cohen M, Chen Y, McCauley M, Gamble T, Hosseinipour MC, Kumarasamy N, HPTN 052 Study Team. Prevention of HIV-1 infection with early antiretroviral therapy. N Engl J Med 2011 Aug 11;365(6):493-505 [FREE Full text] [doi: 10.1056/NEJMoa1105243] [Medline: 21767103]
10. World Health Organization. Summary of 3rd Global Consultation Meeting on HIV Surveillance. Geneva, Switzerland: WHO; 2015 Presented at: Global Consultation Meeting on HIV Surveillance; May 11-14, 2015; Bangkok, Thailand URL: http://www.who.int/hiv/pub/meetingreports/Summary.pdf?ua=1
11. World Health Organization. Consolidated guidelines on person-centred HIV patient monitoring and case surveillance. Geneva, Switzerland: WHO; 2017. URL: http://apps.who.int/iris/bitstream/10665/255702/1/9789241512633-eng.pdf?ua=1 [accessed 2017-12-12] [WebCite Cache ID 6veuHr185]
12. World Health Organization. A Country-Led Platform for Information and Accountability. Geneva, Switzerland: WHO; 2011. Monitoring, Evaluation and Review of National Health Strategies URL: http://www.who.int/healthinfo/country_monitoring_evaluation/1085_IER_131011_web.pdf [accessed 2017-12-12] [WebCite Cache ID 6verWRYHK]
13. World Health Organization. Civil Registration and Vital Statistics 2013: challenges, best practice and design principles for modern systems. Geneva, Switzerland: WHO; 2013. URL: http://www.who.int/healthinfo/civil_registration/crvs_report_2013.pdf [accessed 2017-12-12] [WebCite Cache ID 6vereWQjZ]
14. Low-Beer D, Stoneburner R. AIDS communications through social networks: catalyst for behaviour changes in Uganda. African Journal of AIDS Research 2004 May;3(1):1-13. [doi: 10.2989/16085900409490313] [Medline: 25874978]
15. Medley A, Bachanas P, Grillo M, Hasen N, Amanyeiwe U. Integrating prevention interventions for people living with HIV into care and treatment programs: a systematic review of the evidence. J Acquir Immune Defic Syndr 2015 Apr 15;68 Suppl 3:S286-S296 [FREE Full text] [doi: 10.1097/QAI.0000000000000520] [Medline: 25768868]
16. Joint United Nations Programme on HIV/AIDS. Global AIDS Update. Geneva, Switzerland: UNAIDS; 2016. URL: http://www.unaids.org/sites/default/files/media_asset/global-AIDS-update-2016_en.pdf [accessed 2017-12-12] [WebCite Cache ID 6ves2wODh]
17. Joint United Nations Programme on HIV/AIDS. Finding solutions for everyone at every stage of life. Geneva, Switzerland: UNAIDS; 2016. Get on the Fast Track: The life-cycle approach to HIV URL: http://www.unaids.org/sites/default/files/media_asset/Get-on-the-Fast-Track_en.pdf [accessed 2017-12-12] [WebCite Cache ID 6vesIxjV5]
18. World Health Organization. Global health sector response to HIV, 2000-2015: focus on innovations in Africa. Geneva, Switzerland: WHO; 2015. URL: http://apps.who.int/iris/bitstream/10665/198065/1/9789241509824_eng.pdf [accessed 2017-12-12] [WebCite Cache ID 6vesX4usF]
19. Stoneburner R, Low-Beer D. Population-level HIV declines and behavioral risk avoidance in Uganda. Science 2004 Apr 30;304(5671):714-718 [FREE Full text] [doi: 10.1126/science.1093166] [Medline: 15118157]

Abbreviations

AIDS: acquired immune deficiency syndrome
CDC: Centers for Disease Control and Prevention
MDGs: Millennium Development Goals
M&E: monitoring and evaluation
NGO: non-governmental organization

PEPFAR: US President’s Emergency Plan for AIDS Relief
SDGs: Sustainable Development Goals
TB: tuberculosis
UN: United Nations
UNAIDS: Joint United Nations Programme on HIV/AIDS
WHO: World Health Organization

Edited by P Mee; submitted 08.06.17; peer-reviewed by S Jain, A Pharris, D Stanton; comments to author 13.07.17; revised version received 01.11.17; accepted 13.11.17; published 24.04.18.

Please cite as:
Low-Beer D, Mahy M, Renaud F, Calleja T
Sustainable Monitoring and Surveillance Systems to Improve HIV Programs: Review
JMIR Public Health Surveill 2018;4(2):e3
URL: http://publichealth.jmir.org/2018/2/e3/
doi:10.2196/publichealth.8173
PMID:29691202

Original Paper

Opportunities for Enhanced Strategic Use of Surveys, Medical Records, and Program Data for HIV Surveillance of Key Populations: Scoping Review

Sharon Stucker Weir[1], PhD; Stefan D Baral[2], MD, CCFP, FRCP(C); Jessie K Edwards[3], PhD; Sabrina Zadrozny[4], PhD; James Hargreaves[5], PhD; Jinkou Zhao[6], MD, PhD; Keith Sabin[7], PhD

[1]Carolina Population Center, Department of Epidemiology, University of North Carolina, Chapel Hill, NC, United States
[2]Center for Public Health and Human Rights, Department of Epidemiology, Johns Hopkins School of Public Health, Baltimore, MD, United States
[3]Department of Epidemiology, School of Public Health, University of North Carolina, Chapel Hill, NC, United States
[4]Carolina Population Center, University of North Carolina, Chapel Hill, NC, United States
[5]Department of Social and Environmental Health, London School of Hygiene and Tropical Medicine, London, United Kingdom
[6]Technical.Advice and Partnerships Department, The Global Fund to Fight AIDS, Tuberculosis and Malaria, Geneva, Switzerland
[7]Joint United Nations Programme on HIV/AIDS, Geneva, Switzerland

Corresponding Author:
Sharon Stucker Weir, PhD
Carolina Population Center
Department of Epidemiology
University of North Carolina
Carolina Population Center
123 West Franklin Street
Chapel Hill, NC, 27516
United States
Phone: 1 9194459350
Fax: 1 9194450740
Email: sharon_weir@unc.edu

Abstract

Background: Normative guidelines from the World Health Organization recommend tracking strategic information indicators among key populations. Monitoring progress in the global response to the HIV epidemic uses indicators put forward by the Joint United Nations Programme on HIV/AIDS. These include the 90-90-90 targets that require a realignment of surveillance data, routinely collected program data, and medical record data, which historically have developed separately.

Objective: The aim of this study was to describe current challenges for monitoring HIV-related strategic information indicators among key populations ((men who have sex with men [MSM], people in prisons and other closed settings, people who inject drugs, sex workers, and transgender people) and identify future opportunities to enhance the use of surveillance data, programmatic data, and medical record data to describe the HIV epidemic among key populations and measure the coverage of HIV prevention, care, and treatment programs.

Methods: To provide a historical perspective, we completed a scoping review of the expansion of HIV surveillance among key populations over the past three decades. To describe current efforts, we conducted a review of the literature to identify published examples of SI indicator estimates among key populations. To describe anticipated challenges and future opportunities to improve measurement of strategic information indicators, particularly from routine program and health data, we consulted participants of the Third Global HIV Surveillance Meeting in Bangkok, where the 2015 World Health Organization strategic information guidelines were launched.

Results: There remains suboptimal alignment of surveillance and programmatic data, as well as routinely collected medical records to facilitate the reporting of the 90-90-90 indicators for HIV among key populations. Studies (n=3) with estimates of all three 90-90-90 indicators rely on cross-sectional survey data. Programmatic data and medical record data continue to be insufficiently robust to provide estimates of the 90-90-90 targets for key populations.

Conclusions: Current reliance on more active data collection processes, including key population-specific surveys, remains warranted until the quality and validity of passively collected routine program and medical record data for key populations is optimized.

(JMIR Public Health Surveill 2018;4(2):e28) doi:10.2196/publichealth.8042

KEYWORDS

HIV; sex workers; epidemiology

Introduction

New Vision for Strategic Information

The 2015 World Health Organization (WHO) guidelines for HIV-related strategic information (SI) [1] present a consolidated framework across the full information cycle for HIV (Figure 1), from traditional surveillance indicators, such as HIV prevalence, incidence, and geographic distribution, to indicators monitoring program response, target setting, coverage, and effectiveness. To provide an integrated assessment of progress along one HIV result chain, the guidelines recommend aligning data elements from multiple sources, including health facilities, HIV prevention and treatment programs, and population-based surveys, and disaggregating indicators by population groups. See Figure 1.

Disaggregation of HIV surveillance indicators [2] for the five key populations recognized by WHO (men who have sex with men [MSM], people in prisons and other closed settings, people who inject drugs, sex workers, and transgender people [3]) is an important objective, given that key populations and their sexual partners accounted for about 45% of all new HIV infections in 2015 [4]. Additionally, the role of key populations within the HIV epidemics in countries with generalized epidemics may be greater than previously thought [5,6]. Stigma and social and economic vulnerabilities pose sustained challenges for the uptake of HIV prevention and treatment services among key populations, making it especially important to document gaps in program coverage for these populations [7,8].

Operational Challenges for Indicators for Key Populations

The guidelines recommend estimating key population-specific indicators drawn from multiple data sources; however, the operational challenges of this are significant. First, there are often no standardized definitions for key populations and subgroups. Transgender women are not consistently classified separately from MSM; women who exchange sex for goods or services or work part-time may or may not be included as sex workers; and MSM are often considered one group, despite the significant diversity of HIV acquisition and transmission risks among MSM [9,10]. Eligibility criteria for participation in surveys are generally more restrictive than for programs. Sex worker surveys may restrict participation to women aged 18 years and older who report commercial sex in the past 3 months, whereas programs do not generally screen out part-time sex workers or adolescents who exchange sex for goods or services. An assessment by the Joint United Nations Programme on HIV/AIDS (UNAIDS) and the Global Fund to Fight AIDS, Tuberculosis, and Malaria of the availability and quality of subnational data on sexually transmitted infection (STI) or HIV prevalence, behaviors, and coverage of HIV testing and size estimates for key populations in low- and middle-income countries between 2001 and 2015 found wide variation in the definition of key populations and few examples where definitions were sufficiently consistent to allow trend analysis [11].

Second, as progress is made in reaching the most accessible members of key populations, the HIV epidemic will reach a phase [12] where transmission becomes more concentrated in subgroups that are more hidden, mobile, and harder to reach. Tracking the epidemic among these dynamic subgroups and monitoring program coverage and outreach to them will require an SI system that is flexible, able to identify the emergence of subgroups, and align surveillance trends with program coverage trends. Bio-behavioral surveys such as those recommended by the UNAIDS and WHO Working Group on Global HIV/AIDS and STI Surveillance, currently operate on a 2- to 4-year cycle and, thus, will be challenged to provide flexibility and timely responses to program needs for information. If the coverage area for the survey does not match the catchment area for the specific programs or clinic services, the value of the survey data for immediate local use may be further obscured.

Figure 1. Strategic information cycle.

Third, beyond the challenges of aligning population definitions, identifying subgroups, and geographic reach, there is the challenge of reconciling the different sources of bias in and across survey methods, program monitoring systems, and routine medical record data systems. Different survey and population size estimation methods can result in different HIV prevalence and size estimates for the same population in the same location [13]. Methods such as programmatic mapping rely on information directly obtained from members of the community of interest, as well as indirect information obtained about key populations indirectly from other people who are engaged with the community. This mixed-methods approach may provide information that differs from surveys of population members only. Venue-based estimates are not often adjusted to include individuals who do not attend venues. Additional biases can emerge when sampling does not take mobility of the population into account, or if stigma, fear, or safety concerns limit engagement with key populations at public venues. Estimates from respondent-driven sampling (RDS) may be biased if there is significant clustering by subgroup [14] or geographic area [15]. As more members of key populations are tested and linked to care, it is possible that those who already know they are living with HIV will disproportionately refuse participation in surveys, further biasing the prevalence estimates. Tracking the epidemic and assessing program coverage among persons who move in, out, or across one or more key populations over time [16,17] is challenging but important to monitor coverage among those who may be least likely to be reached. For example, a transgender woman who sells sex and injects drugs may be less likely to access harm reduction services where stigma against transgender women may exist.

Finally, although estimation of the number and percent of new HIV infections is one of the 10 global indicators recommended in the new guidelines, there are methodological and operational challenges of validly measuring incidence among key populations [3]. The most epidemiologically rigorous approach to measure incidence, a prospective cohort study, requires significant resources to develop and maintain, so, it is rarely feasible and can be difficult to interpret given limited generalizability or if the cohort is mobile with significant loss to follow-up [18]. Measuring the social, political, and economic drivers [19,20] of HIV incidence among key populations is also challenging, but efforts to prevent HIV transmission will be limited if we fail to acknowledge the role of these drivers or the structural interventions [20] designed to address them.

We reviewed the published history of HIV surveillance among key populations to put the current challenges into context; identify future opportunities to enhance the use of surveillance data, programmatic data, and medical record data; describe the HIV epidemic among key populations; and measure the coverage of HIV prevention, care, and treatment programs.

Methods

We described the historical expansion of HIV surveillance using published guidelines from WHO and UNAIDS, early surveillance reports identified using MEDLINE, and coauthor recollection.

We described the current salient features of four sources of SI: national household surveys, targeted bio-behavioral surveys, medical record data, and program data sources. We included known strengths, weaknesses, and opportunities for improved use of these data sources, with a focus on estimating elements of the treatment and prevention cascades [21,22].

We also review published estimates of 90-90-90 indicators based on these data sources to identify the availability of cascade estimates for key populations and the extent to which they drew on program, survey, and treatment databases. We conducted a title or abstract search in MEDLINE using the terms: HIV AND [HIV Testing OR population size estimate OR Viral Suppression OR antiretroviral therapy] AND [key populations OR MSM OR sex workers OR injection drug use OR prison OR transgender OR concentrated epidemic] AND [program data OR surveillance data OR routine data OR medical records]. We included manuscripts based on two criteria: (1) it reported any elements of the 90-90-90 HIV treatment cascade (population size, the proportion of the population who know their HIV status, the proportion of the population with HIV that is receiving antiretroviral therapy [ART], and the proportion of the population that is achieving viral suppression) for key populations from low- or middle-income countries and (2) it described strengths or weaknesses of data sources, or issues of aligning data from different sources. For studies where inclusion criteria could not be determined based solely on the abstract, we searched the full article.

Results

History of Surveillance Among Key Populations

Surveillance of the HIV pandemic has evolved over the past three decades, but from the earliest days it has included information about key populations. Early clinical case reports of patients with AIDS in Haiti in 1983 [23], in the Democratic Republic of Congo (Zaire) in 1984 [24], and in Rwanda 1984 [25], focused on clinical manifestations and immunological findings but also noted the presence or absence of behavioral risk, including homosexuality, injecting drug use, and prostitution. In 1985, a human T-cell lymphocyte virus prevalence study in Thailand [26] identified specific high-risk groups as male, homosexual sex workers; thalassemia patients; female sex workers; parenteral drug users; male, venereal disease patients; and blood donors. In 1986, a cohort of gay men in a hepatitis B study was assessed for evidence of AIDS [27]. A *Lancet* report from Rwanda in 1985 conveyed the first clear recommendation to focus on the risk posed by sex work [28]. The report concluded that “Since prostitution is widespread in Central African cities...infection may exist in a large, unconfined group of the general heterosexual population...Among heterosexual populations, prostitutes and probably their male customers should be regarded as high-risk groups.”

Subsequently, guidance surrounding data collection and monitoring for high-risk subgroups started to appear. In 1985, a WHO Coordinating Center Report [29] noted that “An important aspect of WHO activities...will be the collection of data on the incidence of the disease or its causative virus by Member States and the WHO Collaborating Centers...Wherever possible, information on the gender, age, recognized risk factor (if any), and major clinical features should also be provided.” In 1989, the WHO unlinked anonymous testing (UAT) guidelines addressed the ethics of compulsory HIV testing by stating that compulsory testing is unethical and that UAT can only occur when blood was already being taken for another purpose [30]: Information, such as sex and exposure category (if known), may accompany the unlinked anonymous specimen, but the possibility of indirectly identifying people infected with HIV must be eliminated by ensuring that this information is not too discriminating, for example, an age group should be used rather than specific age in years [31]. Despite these guidelines, there is some evidence that sex workers have been subjected to compulsory testing [32,33], at least occasionally since the early years of the HIV epidemic, raising issues of security and confidentiality for key populations.

In 1999, WHO published a comprehensive guide outlining the specific data elements required for STI case reporting in clinical settings [34]. Core elements of this guide included diagnosis, reporting site, date of visit, gender, age group, age, or date of birth. Optional data elements included residence, low education or socioeconomic status, clinical syndrome, anatomic site of infection, date of symptom onset, risk behaviors, pregnancy, history of STI, and treatment. Suggested indicators of risk behavior included the number of sex partners in the past 90 days (or 12 months), whether there were any new sex partners in the past 90 days, the gender of sex partners in the past 12 months (or their sexual orientation), condom use during the last sexual intercourse, drug use in the past 12 months, and giving or receiving money or drugs for sex in the past 12 months.

In 2000, WHO published guidelines for second generation surveillance [35], marking the first description of the strategy to conduct surveillance among antenatal clients in countries with generalized epidemics and among high-risk groups (including sex workers and MSM) among countries with concentrated and low-level epidemics. Second generation surveillance expanded the objectives of surveillance beyond HIV prevalence to include behavioral surveillance and AIDS case reporting. Population-specific questionnaire modules and indicators were developed.

In 2013, these guidelines were updated to incorporate the experiences of countries implementing second generation surveillance over the past 10 years and to incorporate changes in survey methods and laboratory diagnostics [36]. Although guidelines note differing objectives for surveillance (to track how the epidemic in a country is changing) and monitoring and evaluation (to track how effectively programs are responding to the epidemic and whether the outcomes and outputs correspond to the activities planned), the guidelines recommended that the systems be designed to be complementary. Surveillance and survey outcome and impact data should be used to assess the national program response. Program data should provide inputs, outputs, and outcomes to the national monitoring and evaluation system.

In response to each set of guidelines, surveys and surveillance systems were dynamically changed, though the implementation varied by country, region, epidemic profile, and study objective. Currently, many countries have yet to achieve the recommendations from 2103 to align program and surveillance data to describe the HIV epidemic and evaluate the response. Alignment with health record data is a further challenge. Surveillance activities are often implemented with little regard for clinical programs, partly because HIV status and ART were considered too confidential to ask about in surveys. Clinical data can be challenging to triangulate with survey data because definitions of geography or reference period and population do not necessarily align. The current recommendation to estimate the HIV treatment cascade represents a paradigm shift in surveillance and program monitoring because valid estimates of cascade indicators for a district require alignment of definitions, geography, and reference period across survey, program, and treatment databases.

Data From Bio-Behavioral Surveys and Programmatic Mapping

Bio-behavioral surveys of key populations have been the backbone of HIV surveillance for key populations over the past 15 years, particularly for HIV prevalence estimates and more recently for size estimates. They provide probability surveys that facilitate representative estimates. Survey data are used for reporting country-specific indicators to UNAIDS; for use in mathematical models, including spectrum estimates; funding requests; and to guide country-level program reviews. Survey instruments have varied by country but commonly include HIV prevalence, knowledge of HIV transmission routes, sexual behavior (including condom use), and information to estimate the size of populations. Knowledge of HIV status, ART status, and indicators of viral suppression are increasingly included in survey instruments.

Bio-behavioral surveys have strengths relative to health or program data to estimate HIV prevalence and the size of the HIV population. Properly designed and executed bio-behavioral surveys aim to obtain a probability sample of the population in contrast to health sector or program data, which have data only from clients using their services. Surveys can provide an independent evidence-based assessment of gaps in coverage for programs and health facilities and identify emerging epidemics [37,38]. Moreover, validated survey modules on HIV stigma, the accessibility of health services, violence, and sexual behavior can provide more in-depth information on sensitive topics and allow analysis to identify and explore associations between HIV infection and barriers to accessing prevention and other services to guide the implementation of programs. Finally, survey data can provide a profile of those who do not access services, those who are living with HIV but do not know their HIV status, and the profile of those initially linked to treatment who report that they have stopped treatment.

Bio-behavioral surveys, however, have limitations. They are expensive and time-consuming to implement well, including effective engagement of stakeholders. Well-conducted surveys require formative research to guide protocol development, care in translation and back translation, ethical review by the appropriate organizations, interviewer training, ongoing monitoring of data quality, recording of deviations from the protocol, strategies to ensure data confidentiality and protection of participants, and strategies to provide participants with test results and linkage to care if indicated. Because HIV surveillance often takes place outside the health care system and involves contracting an outside implementing institution, there must be careful collaboration between the survey team and those in the health care system to ensure linkage to care for those who test positive as part of the survey.

Although survey design may minimize the effects of selection bias relative to clinical and program data, the effects of self-presentation bias on the validity of self-reported data may be considerable [39,40]. The extent to which anal sex, commercial sex, multiple sexual partnerships, injecting drug use, unprotected sex (sex without a condom), and lack of adherence to ART are underreported is unknown. It is likely that there are shifts in the level of stigma associated with different behaviors that could affect the interpretation of trends from surveillance data. For example, the increased availability of ART has probably led to increased willingness to self-report HIV infection. Legalization of gay marriage may lead to increased willingness to report same sex relationships; crackdowns on MSM may have the reverse effect.

With support from the Global Fund and other donors, countries are using programmatic mapping to identify where to reach key populations and to estimate the size of key populations [41]. Programmatic mapping systematically surveys community informants in a defined geographic area to identify high-risk venues (also known as hotspots) where key populations can be reached. In addition to venue-level data, programmatic mapping can include surveys of a representative sample of venue patrons and workers and oversampling of key populations [42]. Various methods exist for programmatic mapping, but all of them share some strengths and limitations. Strengths for programmatic mapping include the programmatic value of the maps for locating sites where key populations can be reached. Size estimates can be calculated from programmatic mapping and the estimates used to plan outreach visits by peer educators. Limitations for programmatic mapping include the limitations common to other surveys of key populations (as above), as well as the bias arising from the fact that key populations who do not visit these venues and will be missed. Other limitations include the labor-intensive protocol required to ensure that all sites have been listed and a sufficient sample visited.

There are several opportunities for improving the value of surveys for program improvement: (1) aligning size estimates from surveys with program catchment areas; (2) characterizing those reached by the survey, but missed by programs; (3) using the Internet for recruitment of survey participants; (4) measuring the 90-90-90 cascade, including viral suppression; (5) characterizing subgroups [43], including measures of HIV incidence, prevalence, and program coverage; (6) measuring gaps in service delivery; and (7) measuring stigma and its association with access to and use of services. Use of standard stigma indicators in surveys, programs, and health sectors could facilitate improving the quality of care and retention in care and programs.

For example, providing a cluster of differentiation 4 count at the time of the survey in Malawi to female sex workers living with HIV improved the acceptability of HIV testing and facilitated collecting an additional blood sample to estimate the proportion of female sex workers who had achieved viral suppression [44]. Another promising method to identify new HIV infections among key populations is phylodynamic analyses [45,46]. Additional strategies to improve the value of surveys to programs include better engagement of program participants in the design, implementation, and analysis of survey data, including contribution of questions related to specific program elements, as consultants in readiness assessments, as social mobilizers for recruitment of key populations, and as interpreters of the data in data-use workshops**.** In addition, if available, coverage maps from survey data could be provided to programs and workshops held to compare indicators from survey with indicators obtained from programs.

Data From Medical Records

Routinely collected data from medical records and case-based surveillance systems have recently become a focus of development to facilitate measuring progress along the cascade. Where they exist, these data provide the number of persons on treatment. More sophisticated systems monitor progress along the cascade at the individual level from the first positive HIV test to viral suppression. Advocates argue that after HIV diagnosis, all cascade indicators recommended in the new Consolidated Guidelines for HIV Surveillance [47] could be estimated from case-based surveillance [48]. There is the hope that as these systems become routine and new technologies emerge to link data systems, surveillance will improve from the current cross-sectional approach to an ongoing longitudinal dynamic system that can more accurately identify those lost to follow-up.

Challenges in using case-based surveillance and other analyses of medical records for measuring the cascade among key populations in resource-constrained countries are evident already. The system requirements for tracking individual medical records are not available in many settings. In addition to the usual issues associated with solving the problem of deduplicating reports from various sources (eg, interoperability of computer systems, lack of standardization across providers for reporting, lack of timeliness in reporting, and lack of a unique identifier, UID, protocol), a successful case-based surveillance system to monitor the cascade among key populations would require an indicator of key population status in the surveillance record. Inclusion of this indicator is problematic. People may not self-identify as a member of a key population or want the indicator on their medical record [49,50]. Current guidelines do not recommend collecting risk behaviors, that is, key population status, in medical records if it is not clinically relevant. Discriminatory behavior can lead to dangerous situations for members of the populations; in fact, records from MSM services were recently seized by police in Tanzania.

In addition, other factors pose challenges for the unique identification of key population members. It is not unusual for sex workers or gay men to adopt a second identity to hide their affiliation and present at different clinics based on identity. Defining membership in a key population may be differently interpreted across facilities. Membership in a key population may be quite dynamic, causing problems for interpretation of the cascade over time. Biometric UIDs such as fingerprint scans facilitate monitoring at the individual level across data sources but require careful introduction into the community, technical support, data protection schemes, ethical review, and ongoing monitoring.

Due to the stigma associated with being a member of a key population and the lack of a key population identifier in the record, it is likely that case-based surveillance systems will underestimate the size of key populations. Health sector data will overestimate prevalence if those who are infected are more likely to seek clinical care (Berkson’s bias) or if clinics with a higher prevalence of infection among patients are selected for inclusion in surveillance [51], but could underestimate HIV prevalence if clinicians are less likely to directly indicate HIV status in the patient’s medical record (unacceptable disease bias) [52].

Some promising methods, however, are emerging to improve linkages between data sources when UIDs are not available, or an identifier fails to uniquely identify persons. One approach being piloted in the Dominican Republic is a follow-up survey of a sample of persons living with HIV in the treatment database to determine key population membership so that the cascade can be estimated for this subset of persons in the database. Improved probabilistic matching strategies based on available data such as name and birthday may be able to link an individual's records across multiple programs when UIDs are not available. A MEDLINE search in January 2017 of articles related to probabilistic matching of medical records revealed 68 articles on the topic of probabilistic matching of medical records for HIV, of which 24 were published in the past 5 years. Free computer programs to improve deduplication of records are available, and efforts to evaluate the validity of probabilistic matching have been conducted [53].

Finally, new analytic tools are being developed to provide longitudinal measures of the cascade from treatment databases. These cascades indicate the time spent on the pathway from first positive HIV test to reaching viral suppression, the last 90 in the treatment cascade. The longitudinal HIV care and treatment cascade provides an estimate of the person time spent in each of the compartments of the HIV care continuum [54]. In summary, opportunities exist to leverage service delivery data to help both individuals and programs address issues for key populations. There remains much work to improve the quality of these data; survey data will remain of significant value for the foreseeable future; active surveillance from targeted surveys.

Opportunities: Program Data From Key Population Programs

Since the early 1990s, there has been an acknowledgment that key populations are at greater risk of acquiring and transmitting HIV, are less likely to obtain services, and require specific services. Nongovernmental organizations or special outreach programs operating from government clinics may provide more acceptable and tailored services for key populations and may be more willing to engage key populations in target setting, advocacy, and addressing barriers to uptake of services. Some mature programs such as the Avahan sex worker interventions in India [55] collect longitudinal data on at risk populations, conduct size estimates, track intervention coverage, and track HIV prevalence. These programs illustrate that under certain conditions, with adequate resources, leadership, and stable funding, mature programs can set targets based on program data, routinely assess whether targets are met, and only minimally rely on independent HIV surveillance surveys.

Some of the challenges of using program data revolve around the variable quality of program data, arising partially from the broad array of data collection strategies, training, and available support. The main challenge is that program data is not readily generalizable to the entire key population (selection bias), as those who do not visit programs are likely to be different than those who do. Latecomers to programs differ from the early volunteers; thus, the maturity of a program will affect the risk profile of its participants. Even an umbrella program with multiple service delivery sites that employs a UID may comprise an unstable cohort if there is significant mobility across programs and loss to follow-up. Recruitment, attrition, and reach are often not measured systematically within programs even if the capacity to do so theoretically exists [53]. New guidelines for the development, use, and expansion of UIDS for key populations describe some of the technical and ethical challenges in linking program data [56].

National Household Surveys

Probability samples of national household surveys such as the Demographic and Health Surveys (DHS), the AIDS Indicator Surveys, and the newer Population-Based HIV Impact Assessment provide insight into the geographic distribution of HIV across a country but often fail to provide much insight into the HIV epidemic among key populations [57]. Members of key populations may be missed by these surveys for several reasons: they may not be members of a household if they are homeless, resident in brothels or guesthouses, or in school; they may not report [40] sex work, injecting drug use, or male-with-male sex; and they may be more mobile and less likely to be available for an in-home interview. One assessment [58] of the mobility among 1653 female sex workers in Johannesburg, Rustenburg, and Cape Town, South Africa, found that 85% were migrants (39% internal and 46% cross-border). Key populations rarely comprise more than 3% to 8% of the general population. Financial and logistics constraints usually preclude interviewing a sufficient number of people in a household survey to provide an adequate sample of any key population.

Going forward, however, there may be opportunities for greater use of national household survey data. A new incidence model drawing on DHS data estimates the distribution of new infections in a population for groups, including key populations [18]. Estimates of the number of sex workers in the Dominican Republic were also estimated by Bayesian models extrapolating size estimates from known areas to areas without estimates using national survey data available in all areas [59]. The network scale-up method [60,61] has had success with modules added to some household surveys to estimate the size of key populations.

Literature Review: Estimates of the 90-90-90 Treatment Cascade Among Key Populations

We identified 14 publications where at least one of the 90-90-90 indicators was estimated. Three provided estimation of each indicator. The first was an RDS survey among MSM in Moscow, Russia using RDS [62]; the second estimated 90-90-90 indicators using mathematical modeling based on inputs from both survey and programmatic data [63]; and the third was an abstract about MSM and people who inject drugs) in India [64]. Three studies only estimated the first 90 [38,65,66] and four only the second 90 [67-70]. Two studies estimated the first and second 90 [71,72], and three survey-based studies estimated all three indicators of 90-90-90. Among the three studies identified that included estimates of indicators of 90-90-90 from program data, two estimated the second 90 and third 90 [73,74], and one estimated the first 90 and second 90 [75]. In the two studies identified with estimates based in medical record data, only the second 90 was estimated from an ART database [76,77].

Discussion

Principal Findings

Current reliance on active data collection processes, including key population–specific surveillance surveys, is warranted both to collect specific critical information that cannot be obtained from service or other program data and to provide a representative depiction of the HIV epidemic and response. Elements of the latter may be replaced in the future by passively collected routine program and medical record data for key populations. Even in mature programs with years of investment in reaching and treating key populations, national programs must leverage data from program data, medical records, and surveys, as any single source will be insufficient to understand the HIV epidemic, monitor care, and track progress in prevention and along the treatment cascade. Bio-behavioral surveys, although expensive, have proven successful in measuring gaps in program coverage that are not yet revealed by program or medical record data. Improving methods to estimate the treatment cascade from medical records and maintain data security and patient confidentiality will remain a high priority. Strategies to gain insight from multiple sources will require efforts to align geographic catchment areas, definitions, subgroups, and indicators.

Limitations

There are several limitations of this analysis. Important events in the history of surveillance were omitted for brevity. An exhaustive account of the global history of HIV surveillance in the context of key populations is out of scope; however, an overview of the progress and challenges were important to provide some context for the WHO SI guidelines. Important issues regarding measurement of community engagement, data quality, data use, mobility, and the effects of interventions on HIV transmission could not be addressed sufficiently. We did not describe the many size estimation methods available or address issues regarding the validity of these size estimation methods. For transparency, we recommend reporting the methods used to estimate population size and construct denominators for 90-90-90 estimates [11].

Although most peer-reviewed studies identified restricted the analytic sample to highlight a specific key population group, these analyses often included details about additional high-risk behaviors, documenting overlap in population membership. Due to the different risk behaviors of key population subgroups, the overlap between and among groups, and the variety of legal restrictions across countries, the opportunities and challenges related to disaggregating and aligning surveillance data, programmatic data, and medical record data are considerable.

Currently, complete estimates of indicators for the size of key population groups and estimates for indicators of the 90-90-90 targets are not generally available. There has been a movement to increase the use of programmatic data to inform the HIV epidemic among key populations, although evidence of the quality and validity of estimates from these data are lacking [16,78,79,80]. Targeted programs only reach a small fraction of key populations in most countries. Program data often overestimate HIV prevalence and underestimate the size of key populations, possibly because people who are infected are drawn to programs, whereas others avoid it. Size estimates can also be overestimated because of incentives at the program level or individual peer recruiter level if size estimates are larger. Critics suggest that inflated size estimates lead to inflated program targets that are impossible to meet.

Using programmatic data or medical record data for reporting disaggregated estimates of the 90-90-90 targets is particularly challenging when high-risk behaviors are overlapping in key populations, and programs address just one risk behavior. All individuals at risk of HIV acquisition and transmission may not identify as a member of the benefactor population [9,10]. There are programmatic data, particularly from southern and eastern Africa, showing young MSM disproportionately access interventions compared with older MSM. The older MSM are often in relationships with women, and they have heightened fears that their sexual attraction to men may be inadvertently disclosed [81,82]. These older men are uncounted, with high HIV burden, and without targeted services. Collecting robust, high-quality monitoring, evaluation, and surveillance data from programs to estimate the size of populations at highest risk is challenging even when resources are plentiful [55]. Given the challenges in constructing a denominator from programmatic and medical record data, care continua and estimates of 90-90-90 from different data sources are biased or fragmented.

UIDs may offer a solution to linking individuals across surveys, programs, and health care settings if the ethical, logistic, and technological challenges of implementing UIDs for marginalized and criminalized populations can be resolved. In many cases, especially when injection drug use, sex work, or homosexuality is illegal, providing details about high-risk behaviors and linking such information to a permanent medical record is a risk that leaves these already marginalized populations more vulnerable. WHO’s new case reporting and patient monitoring guidelines specifically do not include risk behaviors in the patient monitoring data forms because of the potential for harm to patients from stigmatized populations [83]. These details are therefore often underreported [69,76,84]. Other promising approaches include using the Internet to reach key populations who engage in online community groups.

In conclusion, we recommend ongoing engagement with key population communities in the improvement and alignment of SI indicators across current data sources and exploration of new sources of data. The goal of SI is to improve the adequacy, acceptability, safety, and effectiveness of the public health response to the HIV epidemic among these populations.

Conflicts of Interest

None declared.

References

1. World Health Organization. Geneva, Switzerland; 2015. Consolidated Strategic Information Guidelines for HIV in the Health Sector 2015 URL: http://apps.who.int/iris/bitstream/10665/164716/1/9789241508759_eng.pdf [accessed 2018-02-15] [WebCite Cache ID 6xFtQtr0c]
2. Joint United Nations Programme on HIV/AIDS. 2014. 90-90-90: an ambitious treatment target to help end the AIDS epidemic URL: http://www.unaids.org/sites/default/files/media_asset/90-90-90_en.pdf [accessed 2018-02-15] [WebCite Cache ID 6xFtqxWyQ]
3. World Health Organization. 2016. Consolidated Guidelines On HIV Prevention, Diagnosis, Treatment And Care For Key Populations, 2016 Update URL: http://apps.who.int/iris/bitstream/10665/246200/1/9789241511124-eng.pdf?ua=1 [accessed 2018-02-15] [WebCite Cache ID 6xFu7oBlP]
4. Joint United Nations Programme on HIV/AIDS. 2016. HIV Prevention Among Key Populations URL: http://www.unaids.org/en/resources/presscentre/featurestories/2016/november/20161121_keypops [accessed 2017-05-13] [WebCite Cache ID 6qQxYQzBq]
5. Baral S, Phaswana-Mafuya N. Rewriting the narrative of the epidemiology of HIV in sub-Saharan Africa. SAHARA J 2012;9(3):127-130. [doi: 10.1080/17290376.2012.743787] [Medline: 23237066]
6. Mishra S, Boily MC, Schwartz S, Beyrer C, Blanchard JF, Moses S, et al. Data and methods to characterize the role of sex work and to inform sex work programs in generalized HIV epidemics: evidence to challenge assumptions. Ann Epidemiol 2016;26(8):557-569. [doi: 10.1016/j.annepidem.2016.06.004]
7. Fay H, Baral SD, Trapence G, Motimedi F, Umar E, Iipinge S, et al. Stigma, health care access, and HIV knowledge among men who have sex with men in Malawi, Namibia, and Botswana. AIDS Behav 2011 Aug;15(6):1088-1097. [doi: 10.1007/s10461-010-9861-2] [Medline: 21153432]
8. Risher K, Adams D, Sithole B, Ketende S, Kennedy C, Mnisi Z, et al. Sexual stigma and discrimination as barriers to seeking appropriate healthcare among men who have sex with men in Swaziland. J Int AIDS Soc 2013;16(3 Suppl 2):18715. [doi: 10.7448/IAS.16.3.18715]
9. Evans MG, Cloete A, Zungu N, Simbayi LC. HIV risk among men who have sex with men, women who have sex with women, lesbian, gay, bisexual and transgender populations in South Africa: a mini-review. Open AIDS J 2016;10:49-64. [doi: 10.2174/1874613601610010049]
10. Goldberg S, Zadrozny S, Edwards J, Hileman S, Donastorg Y, Barrington C, et al. HIV Prevention Cascade in the Dominican Republic: Differences by risk behaviors, biological sex, and key populations. 2017 Presented at: Key Population Working Group of the MeSH Consortium; December 7, 2017; Muldersdrift, South Africa.
11. Gall J, Sabin K, Frescura L, Sabin ML, Erkkola T, Toskin I. Global trends of monitoring and data collection on the HIV response among key populations since the 2001 UN declaration of commitment on HIV/AIDS. AIDS Behav 2017 Jul;21(Suppl 1):34-43. [doi: 10.1007/s10461-016-1618-0] [Medline: 27896550]
12. Aral SO, Blanchard JF. Phase specific approaches to the epidemiology and prevention of sexually transmitted diseases. Sex Transm Infect 2002;78(Suppl 1):i1-i2. [doi: 10.1136/sti.78.suppl_1.i1]
13. Weir SS, Merli MG, Li J, Gandhi AD, Neely WW, Edwards JK, et al. A comparison of respondent-driven and venue-based sampling of female sex workers in Liuzhou, China. Sex Transm Infect 2012;88(Suppl 2):i95-101. [doi: 10.1136/sextrans-2012-050638]
14. Goel S, Salganik MJ. Assessing respondent-driven sampling. Proc Natl Acad Sci U S A 2010 Apr;107(15):6743-6747 [FREE Full text] [doi: 10.1073/pnas.1000261107]
15. McCreesh N, Johnston LG, Copas A, Sonnenberg P, Seeley J, Hayes RJ, et al. Evaluation of the role of location and distance in recruitment in respondent-driven sampling. Int J Health Geogr 2011 Oct 18;10:56 [FREE Full text] [doi: 10.1186/1476-072X-10-56] [Medline: 22008416]
16. Granich R, Gupta S, Hall I, Aberle-Grasse J, Hader S, Mermin J. Status and methodology of publicly available national HIV care continua and 90-90-90 targets: a systematic review. PLoS Med 2017 Apr;14(4):e1002253 [FREE Full text] [doi: 10.1371/journal.pmed.1002253] [Medline: 28376085]
17. Granich R, Zuniga JM. Moving from public health failure to opportunity. J Int Assoc Provid AIDS Care 2017;16(2):107-109. [doi: 10.1177/2325957416686197] [Medline: 28088869]
18. Bórquez A, Cori A, Pufall EL, Kasule J, Slaymaker E, Price A, et al. The incidence patterns model to estimate the distribution of new HIV infections in Sub-Saharan Africa: development and validation of a mathematical model. PLoS Med 2016 Sep;13(9):e1002121 [FREE Full text] [doi: 10.1371/journal.pmed.1002121] [Medline: 27622516]
19. Elmes J, Skovdal M, Nhongo K, Ward H, Campbell C, Hallett TB, et al. A reconfiguration of the sex trade: how social and structural changes in eastern Zimbabwe left women involved in sex work and transactional sex more vulnerable. PLoS One 2017;12(2):e0171916 [FREE Full text] [doi: 10.1371/journal.pone.0171916] [Medline: 28225822]
20. Shannon K, Strathdee SA, Goldenberg SM, Duff P, Mwangi P, Rusakova M, et al. Global epidemiology of HIV among female sex workers: influence of structural determinants. Lancet 2015 Jan 3;385(9962):55-71 [FREE Full text] [doi: 10.1016/S0140-6736(14)60931-4] [Medline: 25059947]
21. Hargreaves JR, Delany-Moretlwe S, Hallett TB, Johnson S, Kapiga S, Bhattacharjee P, et al. The HIV prevention cascade: integrating theories of epidemiological, behavioural, and social science into programme design and monitoring. Lancet HIV 2016 Jul;3(7):e318-e322. [doi: 10.1016/S2352-3018(16)30063-7] [Medline: 27365206]

22. Garnett GP, Hallett TB, Takaruza A, Hargreaves J, Rhead R, Warren M, et al. Providing a conceptual framework for HIV prevention cascades and assessing feasibility of empirical measurement with data from east Zimbabwe: a case study. Lancet HIV 2016;3(7):e297-e306. [doi: 10.1016/S2352-3018(16)30039-X]
23. Pape JW, Liautaud B, Thomas F, Mathurin JR, Amand St MM, Boncy M, et al. Characteristics of the acquired immunodeficiency syndrome (AIDS) in Haiti. N Engl J Med 1983;309(16):945-950. [Medline: 6621622]
24. Piot P, Quinn TC, Taelman H, Feinsod FM, Minlangu KB, Wobin O, et al. Acquired immunodeficiency syndrome in a heterosexual population in Zaire. Lancet 1984 Jul 14;2(8394):65-69. [Medline: 6146009]
25. Van de Perre P, Rouvroy D, Lepage P, Bogaerts J, Kestelyn P, Kayihigi J, et al. Acquired immunodeficiency syndrome in Rwanda. Lancet 1984;2(8394):62-65. [Medline: 6146008]
26. Wangroongsarb Y, Weniger BG, Wasi C, Traisupa A, Kunasol P, Rojanapithayakorn W, et al. Prevalence of HTLV-III/LAV antibody in selected populations in Thailand. Southeast Asian J Trop Med Public Health 1985 Dec;16(4):517-520. [Medline: 3869739]
27. Jaffe HW, Hardy AM, Morgan WM, Darrow WW. The acquired immunodeficiency syndrome in gay men. Ann Intern Med 1985 Nov;103(5):662-664. [Medline: 2996396]
28. Van de Perre P, Clumeck N, Carael M, Nzabihimana E, Robert-Guroff M, De Mol P, et al. Female prostitutes: a risk group for infection with human T-cell lymphotropic virus type III. Lancet 1985 Sep 7;2(8454):524-527. [Medline: 2863544]
29. Centers for Disease Control. Acquired immunodeficiency syndrome: meeting of the WHO Collaborating Centres on AIDS. MMWR Morb Mortal Wkly Rep 1985 Nov 08;34(44):678-679. [Medline: 3932833]
30. WHO Global Programme on AIDS. Geneva, Switzerland: World Health Organization; 1989. Unlinked Anonymous Screening for the Public Health Surveillance of HIV Infections: Proposed International Guidelines URL: http://apps.who.int/iris/bitstream/10665/64004/1/GPA_SFI_89.3.pdf [accessed 2018-02-15] [WebCite Cache ID 6xGLr8H9I]
31. Gill ON, Adler MW, Day NE. Monitoring the prevalence of HIV. Br Med J 1989 Nov 25;299(6711):1295-1298 [FREE Full text] [Medline: 2513926]
32. Mpaka C. Ipsnews. 2009. RIGHTS: Police Force HIV Tests for Sex Workers URL: http://www.ipsnews.net/2009/10/rights-police-force-hiv-tests-for-sex-workers/ [accessed 2018-03-15] [WebCite Cache ID 6xvjMkvQ8]
33. Devika J. The unheard scream: reproductive health and women's lives in India. Contrib Indian Sociol 2007;41(3):446-448.
34. UNAIDS/WHO Working Group on Global HIV/AIDS/SI Surveillance. Apps.who.int. Geneva, Switzerland; 1999. Guidelines for sexually transmitted infections surveillance URL: http://apps.who.int/iris/bitstream/10665/66421/1/WHO_CDS_CSR_EDC_99.3.pdf?ua=1 [accessed 2018-02-15] [WebCite Cache ID 6xGNBdkPi]
35. UNAIDS/WHO Working Group on global HIV/AIDS and STI surveillance. Who.int. 2000. Guidelines for second generation HIV surveillance URL: http://www.who.int/hiv/pub/surveillance/en/cds_edc_2000_5.pdf [accessed 2018-02-16] [WebCite Cache ID 6xGamkT6k]
36. UNAIDS/WHO Working Group on Global HIV/AIDS and STI Surveillance. Apps.who.int. 2013. Guidelines for second generation HIV surveillance:an update: know your epidemic URL: http://apps.who.int/iris/bitstream/10665/85511/1/9789241505826_eng.pdf?ua=1 [accessed 2018-02-16] [WebCite Cache ID 6xGb4z1Is]
37. Schaefer R, Gregson S, Takaruza A, Rhead R, Masoka T, Schur N, et al. Spatial patterns of HIV prevalence and service use in East Zimbabwe: implications for future targeting of interventions. J Int AIDS Soc 2017 Dec 28;20(1):21409 [FREE Full text] [doi: 10.7448/IAS.20.1.21409] [Medline: 28364568]
38. Strathdee S, Magis-Rodriguez C, Mays VM, Jimenez R, Patterson TL. The emerging HIV epidemic on the Mexico-U.S. border: an international case study characterizing the role of epidemiology in surveillance and response. Ann Epidemiol 2012 Jun;22(6):426-438 [FREE Full text] [doi: 10.1016/j.annepidem.2012.04.002] [Medline: 22626001]
39. Coughlin SS. Recall bias in epidemiologic studies. J Clin Epidemiol 1990;43(1):87-91. [Medline: 2319285]
40. Catania JA, Gibson DR, Chitwood DD, Coates TJ. Methodological problems in AIDS behavioral research: influences on measurement error and participation bias in studies of sexual behavior. Psychol Bull 1990 Nov;108(3):339-362. [Medline: 2270232]
41. Sabin K, Zhao J, Garcia Calleja JM, Sheng Y, Arias Garcia S, Reinisch A, et al. Availability and quality of size estimations of female sex workers, men who have sex with men, people who inject drugs and transgender women in low- and middle-income countries. PLoS One 2016 May 10;11(5):e0155150 [FREE Full text] [doi: 10.1371/journal.pone.0155150] [Medline: 27163256]
42. Weir SS, Merli MG, Li J, Gandhi AD, Neely WW, Edwards JK, et al. A comparison of respondent-driven and venue-based sampling of female sex workers in Liuzhou, China. Sex Transm Infect 2012 Dec;88 Suppl 2:i95-101 [FREE Full text] [doi: 10.1136/sextrans-2012-050638] [Medline: 23172350]
43. Wilson D. HIV programs for sex workers: lessons and challenges for developing and delivering programs. PLoS Med 2015 Jun;12(6):e1001808 [FREE Full text] [doi: 10.1371/journal.pmed.1001808] [Medline: 26079267]
44. Lancaster KE, Powers KA, Lungu T, Mmodzi P, Hosseinipour MC, Chadwick K, et al. The HIV care continuum among female sex workers: a key population in Lilongwe, Malawi. PLoS One 2016;11(1):e0147662 [FREE Full text] [doi: 10.1371/journal.pone.0147662] [Medline: 26808043]

45. Collins-Fairclough A, Dennis AM, Nelson JA, Weir SS, Figueroa JP. HIV drug resistance surveillance among Jamaican men who have sex with men should be prioritized for reducing HIV transmission. AIDS Res Hum Retroviruses 2015 Aug;31(8):841-844 [FREE Full text] [doi: 10.1089/aid.2015.0040] [Medline: 26133540]
46. Brenner BG, Roger M, Stephens D, Moisi D, Hardy I, Weinberg J, Montreal PHI Cohort Study Group. Transmission clustering drives the onward spread of the HIV epidemic among men who have sex with men in Quebec. J Infect Dis 2011 Oct 01;204(7):1115-1119 [FREE Full text] [doi: 10.1093/infdis/jir468] [Medline: 21881127]
47. World Health Organization. Geneva, Switzerland; 2016. Consolidated Guidelines on HIV Prevention, Diagnosis, Treatment and Care for Key Populations URL: http://apps.who.int/iris/bitstream/10665/246200/1/9789241511124-eng.pdf?ua=1 [accessed 2018-02-16] [WebCite Cache ID 6xGc2Bu6E]
48. Low-Beer NR. Adopting routine patient-level data to track the HIV epidemic and to inform HIV prevention and treatment strategies: a focus on Eastern and Southern Africa. 2016 Presented at: International AIDS Society; 2016; Durban, South Africa URL: http://programme.aids2016.org/Programme/Session/98
49. Decker M, Crago AL, Chu SK, Sherman SG, Seshu MS, Buthelezi K, et al. Human rights violations against sex workers: burden and effect on HIV. Lancet 2015 Jan 10;385(9963):186-199 [FREE Full text] [doi: 10.1016/S0140-6736(14)60800-X] [Medline: 25059943]
50. Islam MM, Topp L, Iversen J, Day C, Conigrave KM, Maher L. Healthcare utilisation and disclosure of injecting drug use among clients of Australia's needle and syringe programs. Aust N Z J Public Health 2013 Apr;37(2):148-154. [doi: 10.1111/1753-6405.12032] [Medline: 23551473]
51. Ng M, Gakidou E, Murray CJ, Lim SS. A comparison of missing data procedures for addressing selection bias in HIV sentinel surveillance data. Popul Health Metr 2013 Jul 24;11(1):12 [FREE Full text] [doi: 10.1186/1478-7954-11-12] [Medline: 23883362]
52. Sackett DL. Bias in analytic research. J Chronic Dis 1979;32(1-2):51-63. [Medline: 447779]
53. Cargo M, Stankov I, Thomas J, Saini M, Rogers P, Mayo-Wilson E, et al. Development, inter-rater reliability and feasibility of a checklist to assess implementation (Ch-IMP) in systematic reviews: the case of provider-based prevention and treatment programs targeting children and youth. BMC Med Res Methodol 2015 Sep 07;15:73 [FREE Full text] [doi: 10.1186/s12874-015-0037-7] [Medline: 26346461]
54. Lesko C, Edwards JK, Moore RD, Lau B. A longitudinal, HIV care continuum: 10-year restricted mean time in each care continuum stage after enrollment in care, by history of IDU. AIDS 2016 Sep 10;30(14):2227-2234 [FREE Full text] [doi: 10.1097/QAD.0000000000001183] [Medline: 27314178]
55. Verma R, Shekhar A, Khobragade S, Adhikary R, George B, Ramesh BM, et al. Scale-up and coverage of Avahan: a large-scale HIV-prevention programme among female sex workers and men who have sex with men in four Indian states. Sex Transm Infect 2010 Feb;86 Suppl 1:i76-i82 [FREE Full text] [doi: 10.1136/sti.2009.039115] [Medline: 20167737]
56. World Health Organization. Apps.who.int. 2016. Uniquely Identifying Cases and Monitoring Patients URL: http://apps.who.int/iris/bitstream/10665/255702/1/9789241512633-eng.pdf [accessed 2018-02-16] [WebCite Cache ID 6xGcu5CMi]
57. Fenton KA, Johnson AM, McManus S, Erens B. Measuring sexual behaviour: methodological challenges in survey research. Sex Transm Infect 2001 Apr;77(2):84-92 [FREE Full text] [Medline: 11287683]
58. Richter M, Chersich MF, Vearey J, Sartorius B, Temmerman M, Luchters S. Migration status, work conditions and health utilization of female sex workers in three South African cities. J Immigr Minor Health 2014 Feb;16(1):7-17 [FREE Full text] [doi: 10.1007/s10903-012-9758-4] [Medline: 23238581]
59. Edwards JK. Quantitative methods using a geographic approach to extrapolate, in Population Size Estimation Consultation: Centers for Disease Control. 2017 Presented at: CDC Conference on Estimation; 2017; Atlanta, Georgia.
60. Bernard HR, Hallett T, Iovita A, Johnsen EC, Lyerla R, McCarty C, et al. Counting hard-to-count populations: the network scale-up method for public health. Sex Transm Infect 2010 Dec;86 Suppl 2:ii11-ii15 [FREE Full text] [doi: 10.1136/sti.2010.044446] [Medline: 21106509]
61. Feehan DM, Umubyeyi A, Mahy M, Hladik W, Salganik MJ. Quantity versus quality: a survey experiment to improve the network scale-up method. Am J Epidemiol 2016 Apr 15;183(8):747-757 [FREE Full text] [doi: 10.1093/aje/kwv287] [Medline: 27015875]
62. Wirtz AL, Zelaya CE, Galai N, Peryshkina A, Moguilnyi V, Dzhigun P, et al. Finding unawares: high levels of undiagnosed HIV infection among Moscow men who have sex with men. 2014 Presented at: CROI; 2014; Boston, MA.
63. Maheu-Giroux M, Vesga JF, Diabaté S, Alary M, Baral S, Diouf D, et al. Population-level impact of an accelerated HIV response plan to reach the UNAIDS 90-90-90 target in Côte d'Ivoire: Insights from mathematical modeling. PLoS Med 2017 Jun;14(6):e1002321 [FREE Full text] [doi: 10.1371/journal.pmed.1002321] [Medline: 28617810]
64. Mehta SH, Solomon S. HIV care cascade among hard to reach populations in India: need to expand hiv counseling and testing. In: Conference on Retroviruses and Opportunistic Infections. 2014 Presented at: Conference on Retroviruses and Opportunistic Infections; 2014; Boston, MA URL: http://www.croiconference.org/sites/default/files/posters/1063_0.pdf
65. Zhang J, Xu JJ, Song W, Pan S, Chu ZX, Hu QH, et al. HIV incidence and care linkage among MSM first-time-testers in Shenyang, China 2012-2014. AIDS Behav 2018;22(3):711-721. [doi: 10.1007/s10461-017-1840-4] [Medline: 28660379]
66. Yi S, Ngin C, Tuot S, Chhoun P, Chhim S, Pal K, et al. HIV prevalence, risky behaviors, and discrimination experiences among transgender women in Cambodia: descriptive findings from a national integrated biological and behavioral survey.

BMC Int Health Hum Rights 2017 May 23;17(1):14 [FREE Full text] [doi: 10.1186/s12914-017-0122-6] [Medline: 28535758]

67. Charurat ME, Emmanuel B, Akolo C, Keshinro B, Nowak RG, Kennedy S, TRUST Study Group. Uptake of treatment as prevention for HIV and continuum of care among HIV-positive men who have sex with men in Nigeria. J Acquir Immune Defic Syndr 2015 Mar 01;68 Suppl 2:S114-S123 [FREE Full text] [doi: 10.1097/QAI.0000000000000439] [Medline: 25723975]
68. Joint United Nations Programme on HIV/AIDS. 2014. UNGASS Country Progress Report, Republic of Armenia URL: http://files.unaids.org/en/dataanalysis/knowyourresponse/countryprogressreports/2014countries/ARM_narrative_report_2014.pdf [accessed 2018-02-16] [WebCite Cache ID 6xGe6WHdJ]
69. McFall A, Mehta SH, Srikrishnan AK, Lucas GM, Vasudevan CK, Celentano DD, et al. Getting to 90: linkage to HIV care among men who have sex with men and people who inject drugs in India. AIDS Care 2016 Oct;28(10):1230-1239 [FREE Full text] [doi: 10.1080/09540121.2016.1168915] [Medline: 27054274]
70. Holland CE, Papworth E, Billong SC, Tamoufe U, LeBreton M, Kamla A, et al. Antiretroviral treatment coverage for men who have sex with men and female sex workers living with HIV in Cameroon. J Acquir Immune Defic Syndr 2015 Mar 01;68 Suppl 2:S232-S240. [doi: 10.1097/QAI.0000000000000443] [Medline: 25723989]
71. Holland CE, Kouanda S, Lougué M, Pitche VP, Schwartz S, Anato S, et al. Using Population-Size Estimation and Cross-sectional Survey Methods to Evaluate HIV Service Coverage Among Key Populations in Burkina Faso and Togo. Public Health Rep 2016 Nov;131(6):773-782 [FREE Full text] [doi: 10.1177/0033354916677237] [Medline: 28123223]
72. Schwartz S, Papworth E, Anato S, Kouanda S, Ky-Zerbo O, Ouedraogo GEA. Low awareness of HIV status and limited uptake of HIV treatment among HIV-infected men who have sex with men in Burkina Faso and Togo. 2014 Presented at: International AIDS Conference; 2014; Melbourne, Australia.
73. Cowan FM, Davey CB, Fearon E, Mushati P, Dirawo J, Cambiano V, et al. The HIV care cascade among female sex workers in Zimbabwe: results of a population-based survey from the Sisters Antiretroviral Therapy Programme for prevention of HIV, an integrated response (SAPPH-IRe) trial. J Acquir Immune Defic Syndr 2017 Apr 01;74(4):375-382. [doi: 10.1097/QAI.0000000000001255] [Medline: 27930599]
74. Donastorg Y, Barrington C, Perez M, Kerrigan D. Abriendo Puertas: baseline findings from an integrated intervention to promote prevention, treatment and care among FSW living with HIV in the Dominican Republic. PLoS One 2014;9(2):e88157 [FREE Full text] [doi: 10.1371/journal.pone.0088157] [Medline: 24551079]
75. Hargreaves JR, Mtetwa S, Davey C, Dirawo J, Chidiya S, Benedikt C, et al. Implementation and Operational Research: Cohort Analysis of Program Data to Estimate HIV Incidence and Uptake of HIV-Related Services Among Female Sex Workers in Zimbabwe, 2009-2014. J Acquir Immune Defic Syndr 2016 May 01;72(1):e1-e8. [doi: 10.1097/QAI.0000000000000920] [Medline: 27093516]
76. National Center for Disease Control and Public Health (NCDCPH) and Country Coordination Mechanism. UNAIDS. 2015. Global AIDS Response Progress Report: GEORGIA Country Progress Report, 2014 URL: http://www.unaids.org/sites/default/files/country/documents/GEO_narrative_report_2015.pdf [accessed 2018-03-15] [WebCite Cache ID 6xvoYoaVz]
77. Joint United Nations Program on HIV/AIDS, Ministry of Health (Bulgaria). Global Health Data Exchange. Geneva, Switzerland: Joint United Nations Program on HIV/AIDS; 2014. Bulgaria Country Progress Report on Monitoring the 2013 Political Declaration On HIV/AIDS, the Dublin Declaration and the Universal Access in the Health Sector Response URL: http://ghdx.healthdata.org/record/bulgaria-country-progress-report-monitoring-2013-political-declaration-hivaids-dublin
78. Joint United Nations Programme on HIV/AIDS. Geneva, Switzerland; 2018. Global AIDS Monitoring 2018: indicators for monitoring the 2016 United Nations Political Declaration on HIV and AIDS URL: http://www.unaids.org/sites/default/files/media_asset/global-aids-monitoring_en.pdf [accessed 2018-02-16] [WebCite Cache ID 6xGaEYaBp]
79. Maleke K, Makhakhe N, Peters RP, Jobson G, De Swardt G, Daniels J, et al. HIV risk and prevention among men who have sex with men in rural South Africa. Afr J AIDS Res 2017 Mar;16(1):31-38. [doi: 10.2989/16085906.2017.1292925] [Medline: 28367747]
80. Hladik W, Benech I, Bateganya M, Hakim AJ. The utility of population-based surveys to describe the continuum of HIV services for key and general populations. Int J STD AIDS 2016 Jan;27(1):5-12. [doi: 10.1177/0956462415581728] [Medline: 25907348]
81. Mayer KH, Bekker LG, Stall R, Grulich AE, Colfax G, Lama JR. Comprehensive clinical care for men who have sex with men: an integrated approach. Lancet 2012 Jul 28;380(9839):378-387 [FREE Full text] [doi: 10.1016/S0140-6736(12)60835-6] [Medline: 22819653]
82. Mantell JE, Tocco JU, Osmand T, Sandfort T, Lane T. Switching on after nine: black gay-identified men's perceptions of sexual identities and partnerships in South African towns. Glob Public Health 2016;11(7-8):953-965 [FREE Full text] [doi: 10.1080/17441692.2016.1142592] [Medline: 26878380]
83. World Health Organization. Geneva, Switzerland; 2017. Consolidated guidelines on person-centred HIV patient monitoring and case surveillance URL: http://apps.who.int/iris/bitstream/10665/255702/1/9789241512633-eng.pdf?ua=1 [accessed 2018-02-16] [WebCite Cache ID 6xGZYVTrt]
84. Gupta S, Granich R. National HIV Care Continua for Key Populations. J Int Assoc Provid AIDS Care 2017;16(2):125-132. [doi: 10.1177/2325957416686195] [Medline: 28090799]

Abbreviations

ART: antiretroviral therapy
DHS: Demographic and Health Surveys
MSM: men who have sex with men
RDS: respondent-driven sampling
SI: strategic information
STI: sexually transmitted infection
UAT: unlinked anonymous testing
UID: unique identifier
UNAIDS: Joint United Nations Programme on HIV/AIDS
WHO: World Health Organization

Edited by P Mee; submitted 14.05.17; peer-reviewed by V Hope, A Bourne, W Hladik, C Conn, A Sharma; comments to author 13.07.17; revised version received 14.11.17; accepted 20.12.17; published 22.05.18.

Please cite as:
Weir SS, Baral SD, Edwards JK, Zadrozny S, Hargreaves J, Zhao J, Sabin K
Opportunities for Enhanced Strategic Use of Surveys, Medical Records, and Program Data for HIV Surveillance of Key Populations: Scoping Review
JMIR Public Health Surveill 2018;4(2):e28
URL: http://publichealth.jmir.org/2018/2/e28/
doi:10.2196/publichealth.8042
PMID:29789279

Viewpoint

Studying Cannabis Use Behaviors With Facebook and Web Surveys: Methods and Insights

Jacob T Borodovsky[1*], BA; Lisa A Marsch[1*], PhD; Alan J Budney[1*], PhD

The Center for Technology and Behavioral Health, Dartmouth Geisel School of Medicine, Lebanon, NH, United States

[*]all authors contributed equally

Corresponding Author:
Jacob T Borodovsky, BA
The Center for Technology and Behavioral Health
Dartmouth Geisel School of Medicine
46 Centerra Parkway
Lebanon, NH, 03766
United States
Phone: 1 603 646 7004
Fax: 1 603 646 7068
Email: jacob.borodovsky@gmail.com

Abstract

The rapid and wide-reaching expansion of internet access and digital technologies offers epidemiologists numerous opportunities to study health behaviors. One particularly promising new data collection strategy is the use of Facebook's advertising platform in conjunction with Web-based surveys. Our research team at the Center for Technology and Behavioral Health has used this quick and cost-efficient method to recruit large samples and address unique scientific questions related to cannabis use. In conducting this research, we have gleaned several insights for using this sampling method effectively and have begun to document the characteristics of the resulting data. We believe this information could be useful to other researchers attempting to study cannabis use or, potentially, other health behaviors. The first aim of this paper is to describe case examples of procedures for using Facebook as a survey sampling method for studying cannabis use. We then present several distinctive features of the data produced using this method. Finally, we discuss the utility of this sampling method for addressing specific types of epidemiological research questions. Overall, we believe that sampling with Facebook advertisements and Web surveys is best conceptualized as a targeted, nonprobability-based method for oversampling cannabis users across the United States.

(JMIR Public Health Surveill 2018;4(2):e48) doi:10.2196/publichealth.9408

KEYWORDS

epidemiology; cross-sectional studies; sampling studies; social media; data collection; cannabis; surveys

Introduction

Surveys have been a methodological cornerstone of epidemiology since the inception of the field. However, the manner in which epidemiologists conduct surveys has undergone several paradigm shifts in tandem with advances in mass communication and information dissemination [1]. Initially, data collection was limited to local in-person site visits [1,2]. Over time, new modes of communication, such as mailed questionnaires and random digital dialing [3], expanded epidemiologists' methodological toolkit [1]. Once the internet began to function as a principal means of communication, it too was recognized for its potential utility as a data collection method [1,4,5]—particularly for substance use data [6]. Historically, collecting data on hidden and stigmatized populations such as substance users had been exceptionally difficult. Early pioneering work demonstrated that it was possible to use the internet to study these populations [7,8].

Today, approximately 3.5 billion people around the world have internet access [9], and 2.3 billion people own a smartphone [10], allowing them to access the internet at any time of day in almost any location. Approximately 2.5 billion people [11] now use social media and networking sites such as Facebook and Twitter, primarily on their smartphones [12,13], for an average of 1 to 2 hours per day [14,15]. In parallel with this increase in internet and social media use, researchers have devoted considerable attention to conducting Web-based studies of health behaviors. In doing so, they have developed various Web survey- and social media-based data collection methods [16-24]. Of the social media platforms now commonly used for health research purposes, Facebook, in particular, has emerged as a useful and low-cost means of recruiting participants [25] from hard-to-reach populations [16]. At present, researchers have used Facebook to disseminate Web surveys to study a range of diseases and health behaviors such as HIV [26]; vaccine uptake [27,28]; mammographies [29]; contraception [30]; several mental health disorders [31-33]; prescription medication misuse [34]; and use of alcohol [35], tobacco [36], e-cigarettes [37], and cannabis [38-42]. Such studies are part of a growing subdiscipline of epidemiology—often termed *Infodemiology* [43], *Digital Epidemiology* [44], or *E-Epidemiology* [45]—that is characterized by the overlap between traditional epidemiological research goals and the utilization of new digital infrastructures.

Facebook-based Web surveys are well suited for addressing questions that arise from sociocultural changes because they allow for the rapid study of behaviors on a population-level scale. One example of such a sociocultural change is the shifting legal landscape of cannabis in the United States. More recreational and medical legal cannabis laws (LCL) have been enacted since 2010 than were enacted from 1996 to 2009 (LCL is used here to differentiate such laws from those focused on criminalizing and prohibiting cannabis) [46], and several new concerns—including the expansion of cannabis product diversity [47-50]—have emerged as a result. Many of the questions brought about by cannabis legalization represent excellent scientific targets for Facebook-based sampling methods.

At the Center for Technology and Behavioral Health at Dartmouth College, we have been leveraging the Facebook Business advertising platform to conduct a series of Web surveys to study questions concerning patterns of use of new methods of cannabis administration (eg, vaping and edibles) and relationships between cannabis use and psychological constructs such as anxiety, pain, and readiness to reduce or stop cannabis use. In the process of conducting these surveys, we have learned several lessons about advertising procedures that minimize costs and maximize survey participation. In addition, given the need to determine strengths and limitations of social media-based data collection [51], we have aggregated and examined the data from several of our surveys to identify the unique characteristics of our sampling method.

This paper provides an overview of our procedures and lessons learned using Facebook advertisements and Web surveys as a method to study cannabis use, our findings regarding the unique characteristics of the cannabis use data produced by this method, and how the characteristics of the resulting data clarify the types of research questions best suited for study with this sampling method.

Using the Facebook Advertising Platform

Targeting Parameters

The Facebook advertising platform provides researchers with access to a large sample pool and a wide range of demographic, behavioral, and psychographic targeting parameters. These parameters can be tailored to send customized advertisements (eg, recruitment messages for a Web survey) to the phone and computer screens of specific populations of interest [31,52,53]. This sampling capability is possible because of the vast amount of information Facebook collects about each user's Web-based behavior.

General targeting parameter categories include any age range above 13 years, gender, education (eg, type of educational degree), relationship status (eg, marital status), financial status (eg, income level), geographic location, multicultural affinity (eg, African American, Hispanic, Asian), generation (eg, baby boomers), employment (eg, job title), household composition (eg, new parents), and Web-based purchase behaviors (eg, pain relief medications, alcoholic beverages).

Our research group has been interested in studying cannabis use. However, Facebook does not provide targeting parameter categories such as *cannabis user*. Thus, in our advertising strategy, we use targeting parameters that we believe are correlated with our behavior of interest (cannabis use), such as notable individuals associated with cannabis use (eg, Bob Marley, Ed Rosenthal), cannabis-related magazines (eg, Cannabis Culture, High Times), organizations (eg, Americans for Safe Access, NORML, Weedmaps), and behaviors or topics (eg, "smoking weed","legalize marijuana").

Algorithm Learning and Optimization

Facebook's machine learning–based algorithms are designed to present users with content (including advertisements) relevant to their personality or lifestyle [54-56]. However, the algorithms used to distribute advertisements appear to require a sufficient amount of time to complete a learning phase before they become effective. During this learning phase, enough data must be accumulated to determine which members of the target population have the highest probability of engaging with the advertisement [54-56]. Although we cannot verify that such learning processes are taking place or how they are programmed, our experiences using the Facebook advertisement system to date have been congruent with this literature. Currently, when we use Facebook advertisements to reach cannabis users, we begin advertising with a low spending limit of US $10 per day for 48 to 72 hours, which we believe facilitates algorithm learning. After this 48- to 72-hour time frame, we have consistently seen a notable increase in the rate of clicks. At that point, we have increased the amount of money spent to anywhere between US $30 and US $60 per day. Figure 1 displays the Facebook advertisement click results from our most recently published survey [57]. In this study, we preprogrammed advertisements to be displayed for 6 days (September 3-8, 2016). As is evident from Figure 1, we received few advertisement clicks during the first 24 hours of advertising (September 3). However, after 72 hours, the advertisement was receiving over 500 clicks per day. Of note, Facebook paces advertisement spending to remain within-budget over the course of the entire advertising time frame [58]. We believe this is why the number of clicks per day begins to decline toward the end of the advertising time frame (ie, September 8 in Figure 1).

It is important to note that this specific 48- to 72-hour time frame may not apply to other social media advertising platforms or even to Facebook in the future. However, we believe that the principle for effectively using machine learning–based advertising demonstrated here will still hold.

Figure 1. Click and cost results from a recent cannabis-targeted Facebook advertising campaign.

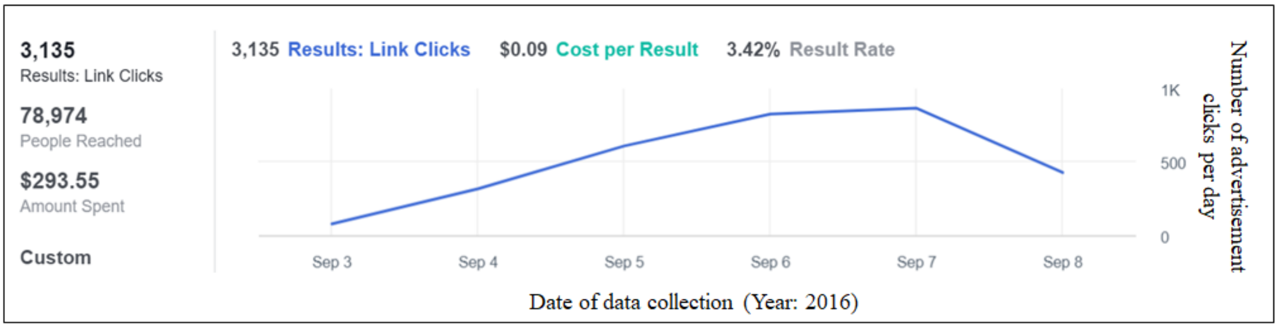
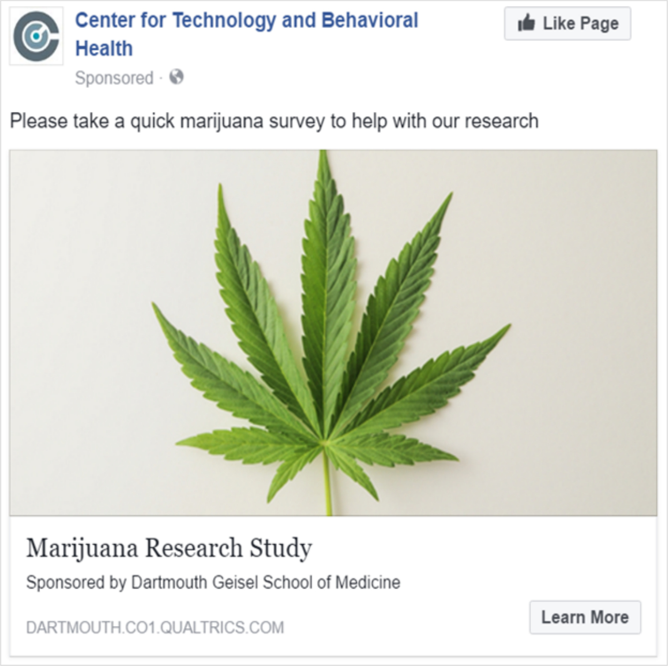


Figure 2. Sample advertisement distributed via Facebook advertising.



Please take a quick marijuana survey to help with our research

Marijuana Research Study

Sponsored by Dartmouth Geisel School of Medicine

DARTMOUTH.CO1.QUALTRICS.COM

Advertisement Imagery and Data Collection

To sample cannabis users in these studies, we direct Facebook recruitment advertisements (Figure 2) to individuals who live in the United States and are affiliated with cannabis-related targeting parameters discussed previously. In the advertisements, we use images and wording that are salient to the population we want to study. This means our advertisements frequently display cannabis leaves and use the words *cannabis* or *marijuana* in the text. The advertisements contain the URL link to our anonymous Qualtrics-hosted survey. Individuals who click the advertisement are redirected to the consent page of the survey. Individuals are excluded if they (1) do not provide consent, (2) do not meet study-specific age requirements (eg, aged above 18 years), or (3) self-report never having used cannabis. We use internal data checks to confirm the veracity of the data. For example, individuals who report their age of initiation of cannabis as being older than their current age are excluded from analyses. We disable internet protocol (IP) address collection to maintain participant anonymity and describe this procedure and its implications clearly on the consent page. We enable Qualtrics data quality features that use cookies to prevent individuals from responding multiple times. We also use the captcha verification feature to prevent responses from internet bots. We have not used any survey completion compensation or incentives in these studies.

Iterative Questionnaire Modifications

In addition to using the sampling strategies discussed previously, we have also iteratively modified the structure of our questionnaires in several ways. Note that our team has not attempted to evaluate the isolated impact of each of these modifications systematically, and thus we cannot comment on causal relationships. However, based on our experiences to date, we believe that the following questionnaire modifications have helped increase the likelihood of participation and completion: (1) using language at the top of the consent page that highlights the changing landscape of US cannabis legalization and the need for public contribution to better understand cannabis; (2) using the term *cannabis* rather than *marijuana*; (3) using patently objective language about cannabis on the consent page (ie, explaining that our research team's primary aim is to collect accurate data—not promote or demonize cannabis); (4) conveying to participants, at the top of the consent page, that the questionnaire will only take 10 to 15 min to complete; (5) conveying to participants that their IP addresses will not be tracked; (6) shortening the overall length of the questionnaire; (7) using images of cannabis and methods of cannabis administration throughout the questionnaire; (8) ensuring that the first few questions that participants see immediately after providing consent are interesting to them (ie, “Have you ever used cannabis?”); (9) distributing uninteresting questions (eg, demographic questions) throughout the questionnaire; and (10) providing an open-ended, free-response item that asks participants for their thoughts about both positive and negative experiences with cannabis.

We believe that our surveys have become increasingly well received by the Facebook community as we have made these changes over time. For example, in our second survey—conducted over 28 days for US $809—our advertisements received 107 *likes* (ie, a positive reaction to the advertisement), 32 comments, and 27 *shares* (ie, an individual sending the advertisement to one of their friends via Facebook). In our fourth survey—conducted over 6 days for US $293—we received 354 likes, 41 comments, and 139 shares. Content analysis of the reactions and comments to the various surveys over time is beyond the scope of this paper, but we believe that the patterns observed generally show increased acceptance and willingness to engage and participate in these types of survey studies.

Cannabis-Related Findings From Facebook Advertising–Generated Survey Data

We have conducted a series of Web surveys using these methods, which have provided insights into patterns of cannabis use across the United States. Below we present data from 6 of those surveys, 3 of which have been published [57,59,60]. Table 1 provides an overview of the recruitment results for each of the surveys.

In our initial study, we used Facebook advertising to recruit adults (N=2910) for a survey on cannabis vaping and edible use. We found that over half (61.27%, 1783/2910) of the users had tried vaping, but only 12.44% (362/2910) of the sample preferred vaping over other methods [61], and the likelihood of vaping and edible use was positively associated with the number of within-state cannabis dispensaries per capita [59]. A second study replicated these findings with cannabis users (N=933). Again, the majority (55.2%, 515/933) had tried vaping, and of those, 27.2% (140/515) had vaped while driving, 34.2% (176/515) had vaped in public, and 14.2% (73/515) had vaped at work (A J Budney, unpublished data, July 2016). In a third study, we collected cannabis use survey data from younger cannabis users (aged 14-18 years; N=2630) and replicated our previous adult findings. We observed the same relationship between dispensaries and vaping and edible use. We also found that home cultivation provisions of legal cannabis laws were uniquely related to a greater likelihood of having used edibles [60]. In a follow-up survey of adults, we collected responses (N=1813) to explore the relationship between provisions of legal cannabis laws (home cultivation and dispensaries) and cannabis edible procurement behaviors. We determined that those who live in states that permit home cultivation tend to make their own cannabis edibles, whereas those who live in states with cannabis dispensaries primarily purchase their edibles [57]. In another study, cannabis users (N=1212) participated in a survey assessing the frequency of cannabis use and thoughts about decreasing their use. A sizeable portion (19.39%, 235/1212) reported having “been concerned about their cannabis use,” and, among these individuals, 75.32% (177/235) had recently thought about reducing their use (A J Budney, unpublished data, February 2017). In another study, cannabis users (n=3561) participated in a survey on frequency of cannabis use and selected psychological processes. Indirect effects of anxiety sensitivity on past 30-day heavy cannabis use were mediated by coping-related motives for cannabis use (A A Knapp, unpublished data, November 2016).

This Facebook-based Web survey approach can be particularly useful for expeditiously conducting studies that help clarify observations and questions that arise from prior survey studies. For example, the primary analyses of the youth survey discussed previously revealed that state-level permission of cannabis home cultivation was statistically related *only* to an elevated likelihood of edible use, whereas state-level permission of cannabis dispensaries was related to elevated likelihoods of *both* lifetime vaping and edible use. After developing a hypothesis to explain this observation, we conducted a new survey 3 months later to better understand these relationships. These data replicated the results from the prior study and further demonstrated that those who grow cannabis were more likely to make edibles at home by economizing low-tetrahydrocannabinol (THC) leftover parts of the cannabis plant. However, those who lived in states with dispensaries were more likely to purchase cannabis edibles [57].

Table 1. Recruitment results from 6 cannabis use Web surveys disseminated using Facebook advertisements (ads).

Survey[a]	Ad delivery time frame	Total ad cost ($ USD)	No. of people who saw ads	No. of ad clicks	Sample size[b]	Ad images	Demographics and cannabis use patterns of recruited sample			
							Age, mean (SD)	Male, n (%)	White, n (%)	Current use[c], n (%)
Survey 1[d]	43 days	800	168,894	3708	2838	Cannabis leaf; College logo	32 (16)	2391 (84.24)	2048 (72.16)	2333 (82.20)
Survey 2	28 days	809	231,400	3932	933	Multiple[e]	44 (18)	758 (81.2)	794 (85.1)	724 (77.6)
Survey 3	20 days	350	126,945	5480	2630	Cannabis leaf	16 (1)	1201 (45.67)	2067 (78.60)	2185 (83.08)
Survey 4	6 days	293	78,974	3135	1813	Cannabis leaf	48 (13)	1386 (76.48)	1608 (88.70)	1540 (85.94)
Survey 5[d]	9 days	402	68,525	2599	1212	Cannabis leaf	28 (11)	784 (64.69)	1029 (84.90)	1132 (93.40)
Survey 6[d]	7 days	377	96,096	5612	2972	Cannabis leaf; Cannabis plant	35 (10)	1815 (61.07)	2653 (89.27)	2549 (85.77)

[a] Eligibility criteria for all surveys: (1) lifetime cannabis user, (2) age 18 years or older, (3) provided consent/assent, and (4) currently living in the United States. Exception for survey 3 in which the age eligibility criteria was 14 to 18 years.

[b] Respondents who (1) met eligibility criteria, (2) passed data-quality checks, and (3) completed the survey.

[c] “Current use” indicates individuals who used cannabis at least once in the past 30 days.

[d] Participants permitted to skip questions. Reported sample sizes may vary depending on the variable analyzed.

[e] Advertisement images included cannabis leaf, cannabis plant material, Dartmouth College logo, methods of use (eg, joints), depictions of smoking behavior, and smoke clouds.

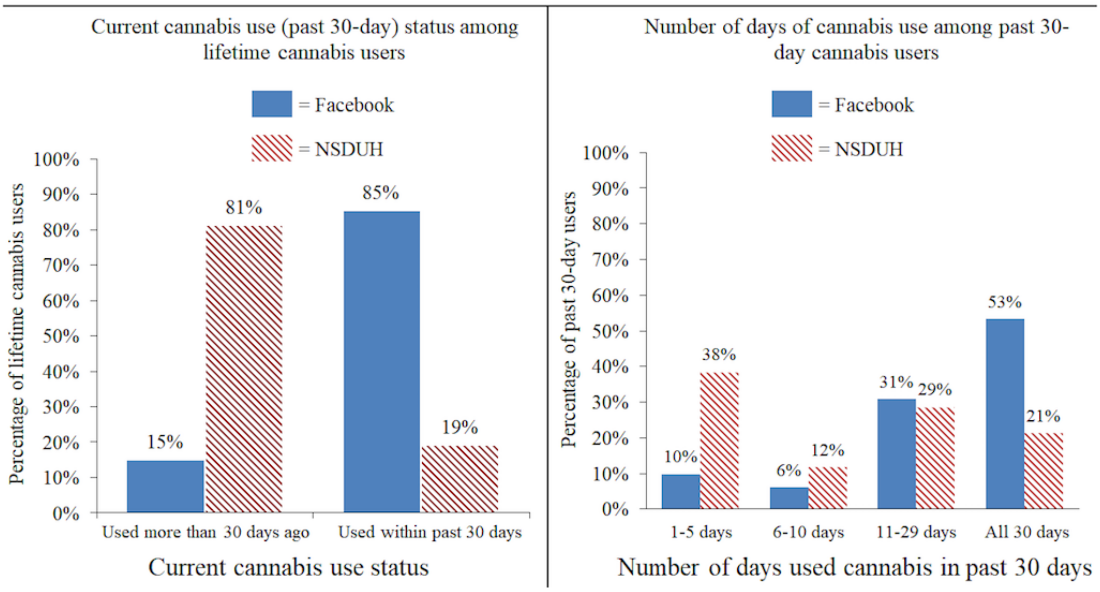
Determining and Interpreting the Idiosyncrasies of Facebook Advertising–Based Cannabis Use Data

As this line of research expands, it is crucial to understand both the strengths and limitations of this Facebook survey sampling method. Epidemiologists use a variety of sampling methods to answer different types of research questions [62-65]. For example, accurately estimating the prevalence of lifetime cannabis use in the United States requires probability-based methods, but understanding the unique cannabis use disorder treatment needs of various population subgroups requires nonprobability-based sampling methods [66]. The unique properties of Facebook sampling for studying cannabis use are largely unclear at this point — making it difficult to determine which types of research questions are most effectively addressed with this method. To begin to fill this gap, we have conducted several secondary analyses of our Facebook-based data. On the basis of our understanding that Facebook advertising was originally designed to help businesses contact their specific niche audience, we hypothesized that Facebook advertisements would produce data similar to traditional nonrandom epidemiological sampling methods, thus producing data on a relatively homogenous subpopulation.

Patterns of Cannabis Use

In our initial surveys, we were interested in studying patterns of use of novel methods of cannabis administration. We assumed that current and regular cannabis users would be the most likely to have used these methods of administration, and therefore, we attempted to oversample this subgroup by using the cannabis-centric sampling targets (eg, Medical Marijuana or High Times Magazine) discussed previously. To confirm that our sampling method did indeed oversample this subgroup of cannabis users, we compared our data with cannabis use data from the probability-based National Survey on Drug Use and Health (NSDUH). The NSDUH can be used to generate prevalence estimates of multiple subtypes of cannabis users in the United States. Figure 3 displays our Facebook-based data combined across several of our published and unpublished adult (aged above 18 years) surveys (N=10,427; includes individuals who did not complete an entire survey) in relation to data from the 2015 NSDUH [67]. The left panel of Figure 3 shows that samples of lifetime cannabis users from our surveys comprised proportionally more current (ie, used at least one time in the past 30 days) users compared with lifetime users in the NSDUH. In addition, the right panel in Figure 3 shows that our samples of current users (n=8886) comprised proportionally more daily cannabis users compared with current users in the NSDUH [59,60,68]. These findings support the notion that Facebook sampling can effectively obtain data from subgroups of lifetime cannabis users who are currently using cannabis on a regular basis.

Figure 3. Comparisons between Facebook-generated cannabis use data and cannabis use data from the 2015 National Survey on Drug Use and Health (NSDUH).



Similar comparative analyses by other researchers support this finding. For example, Barratt et al analyzed cannabis use data from the Global Drug Survey (a self-selected Web survey) in relation to several national probability-based surveys (including NSDUH). Their results demonstrate that Web-based purposive recruitment (including the use of Facebook) can generate samples in which current cannabis users are over-represented [69].

Geographic Representation

A second consideration is the possibility of a selection bias induced by state residence. Individuals living in medical or recreational LCL states may have a greater (or lesser) propensity to take our surveys than individuals from states that have not legalized cannabis use. To determine if this was occurring, we compared several of our Facebook-based datasets (both published and unpublished) with US census data [70]. As demonstrated in Figure 4, Facebook consistently generates samples in which the proportion of survey respondents from each US state matches the proportion of the total US population represented in each US state. Pearson and Spearman correlation coefficients for each relationship displayed in Figure 4 range from .82 to .95 (P<.001). Thus, the results from these data are not necessarily biased by disproportionate geographic representation.

These two characteristics of cannabis use data collected using Facebook advertisements and Web surveys provide some indication as to the types of research questions this method can be useful for studying. Using targeted interests (eg, Medical Marijuana) to recruit participants introduces a selection bias, and our comparisons to NSDUH data suggest that this selection bias generates samples that over-represent current and regular cannabis users. However, our comparisons to the US census indicate that these data are not provincial—Facebook samples individuals from across the United States. Thus, this sampling method seems better suited for research questions aimed at understanding how cannabis legalization may affect current and frequent cannabis users rather than how it will affect inexperienced or infrequent cannabis users.

A strength of this sampling method is that researchers can collect enough data to address their research question after only a few days of data collection. This advantage, however, begs questions related to the potential impact of natural temporal variability on the data and results. For example, how similar are data sampled at the beginning of the week to data sampled at the end of the week? Without answers to this type of question, it is difficult to know the extent to which we should be concerned about temporally related confounding. In our analyses, we have begun preliminary exploration of potential fluctuations in the types of cannabis users who take our surveys on particular days of the week. Table 2 displays demographic and cannabis use characteristics of participants from our recently published youth dataset [60] aggregated by the day of the week that the data were sampled. The resulting profile is relatively stable across days of the week for many variables, but significant variability does occur. Looking at the gender variable in Table 2, it is clear that the distribution of responses across categories within a variable can change multiple times during the week. In addition, abrupt changes in the absolute difference in proportions between categories of a variable can occur (eg, the difference between male and female representation on Saturdays). Given that the sampling strategy remained the same for the duration of the study, it is unclear why such variability occurred and why some variables—such as gender in this particular instance—might be affected more than others. Due to the current degree of uncertainty surrounding this methodological issue, we suggest collecting data for at least a full week to capture any potential temporally related variability in the data. We also suggest that relationships among sample size, effect size, number of statistical tests conducted, and P values be considered as a whole to determine whether the observed fluctuations warrant statistical adjustment.

Additional Considerations for Facebook Sampling

Several other methodological considerations warrant comment. Like other sampling methods, Facebook advertising is subject to both noncoverage and nonresponse biases [71]. There are cannabis users living in the United States who do not have Facebook or have Facebook but do not engage in cannabis-related activity on Facebook (and thus cannot be targeted by our advertising). Other cannabis users may not wish to engage in research studies even if they are presented with the opportunity to do so (ie, they see the advertisement on their screen but do not click it). Although it is difficult to test for systematic differences between those who do and do not click on an advertisement when given a chance, the cannabis users who do click advertisements and take surveys provide generally reliable and diagnostically valid cannabis use data [72].

Figure 4. Percentage of US population in each state versus percentage of Facebook sample in each state.

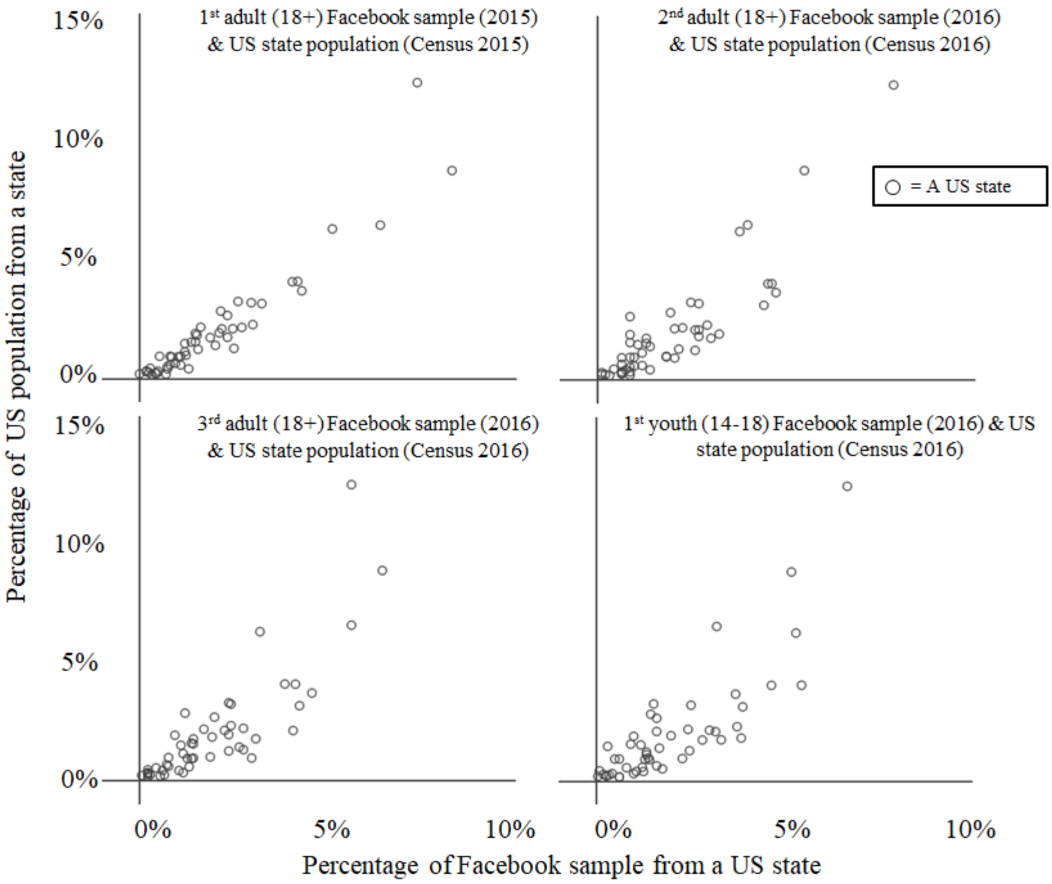


Table 2. Demographic and cannabis use–related data sampled via Facebook advertising according to the day of the week data were collected.

Variable	Day of the week data were collected (N=2630)							*P* value[a]
	Monday[b] (n=435)	Tuesday[b] (n=408)	Wednesday[b] (n=342)	Thursday[c] (n=257)	Friday[b] (n=417)	Saturday[b] (n=365)	Sunday[b] (n=406)	
Education, n (%)[d]								.11
6th grade	0	0	0	0	1 (0.2)	2 (0.5)	1 (0.2)	
7th grade	8 (1.8)	7 (1.7)	3 (0.9)	4 (1.6)	6 (1.4)	5 (1.4)	2 (0.5)	
8th grade	34 (7.8)	31 (7.6)	39 (11.4)	16 (6.2)	48 (11.5)	46 (12.6)	43 (10.6)	
9th grade	89 (20.5)	75 (18.4)	67 (19.6)	51 (19.8)	84 (20.1)	89 (24.4)	87 (21.4)	
10th grade	135 (31.0)	112 (27.5)	102 (29.8)	73 (28.4)	107 (25.7)	100 (27.4)	109 (26.8)	
11th grade	108 (24.8)	108 (26.5)	68 (19.9)	71 (27.6)	111 (26.6)	81 (22.2)	110 (27.1)	
12th grade	40 (9.2)	50 (12.3)	41 (12.0)	27 (10.5)	45 (10.8)	32 (8.8)	44 (10.8)	
Started college	21 (4.8)	25 (6.1)	22 (6.4)	15 (5.8)	15 (3.6)	10 (2.7)	10 (2.5)	
Race, n (%)[d]								.38
African American	24 (5.5)	12 (2.9)	12 (3.5)	8 (3.1)	9 (2.2)	11 (3.0)	13 (3.2)	
Native American	13 (3.0)	15 (3.7)	7 (2.0)	9 (3.5)	16 (3.8)	11 (3.0)	7 (1.7)	
Asian	3 (0.7)	5 (1.2)	5 (1.5)	4 (1.6)	7 (1.7)	3 (0.8)	5 (1.2)	
White	332 (76.3)	311 (76.2)	274 (80.1)	209 (81.3)	341 (81.8)	281 (77.0)	319 (78.6)	
Pacific Islander	2 (0.5)	3 (0.7)	0 (0.0)	1 (0.4)	1 (0.2)	1 (0.3)	1 (0.2)	
Hispanic	61 (14.0)	62 (15.2)	44 (12.9)	26 (10.1)	43 (10.3)	58 (15.9)	61 (15.0)	
Gender, n (%)								<.001
Male	182 (41.8)	197 (48.3)	178 (52.0)	150 (58.4)	191 (45.8)	129 (35.3)	174 (42.9)	
Female	233 (53.6)	200 (49.0)	155 (45.3)	99 (38.5)	209 (50.1)	223 (61.1)	218 (53.7)	
Other	20 (4.6)	11 (2.7)	9 (2.6)	8 (3.1)	17 (4.1)	13 (3.6)	14 (3.4)	
Living situation, n (%)								.29
Lives with neither parent	49 (11.3)	43 (10.5)	37 (10.8)	29 (11.3)	44 (10.6)	32 (8.8)	35 (8.6)	
Lives with both parents	185 (42.5)	187 (45.8)	146 (42.7)	129 (50.2)	176 (42.2)	137 (37.5)	181 (44.6)	
Lives with mother only	153 (35.2)	132 (32.4)	124 (36.3)	74 (28.8)	155 (37.2)	158 (43.3)	147 (36.2)	
Lives with father only	48 (11.0)	46 (11.3)	35 (10.2)	25 (9.7)	42 (10.1)	38 (10.4)	43 (10.6)	
Preferred method, n (%)								.34
Smoke cannabis	231 (83.4)	236 (81.1)	207 (85.2)	162 (85.3)	231 (83.7)	202 (86.7)	229 (84.8)	
Vaporize cannabis	22 (7.9)	21 (7.2)	17 (7.0)	17 (8.9)	17 (6.2)	11 (4.7)	11 (4.1)	
Eat cannabis	24 (8.7)	34 (11.7)	19 (7.8)	11 (5.8)	28 (10.1)	20 (8.6)	30 (11.1)	
Lifetime days smoked cannabis, n (%)[e]								.86
Once	10 (2.3)	10 (2.5)	6 (1.8)	4 (1.6)	10 (2.4)	10 (2.7)	7 (1.7)	
2-5 days	28 (6.5)	24 (5.9)	24 (7.1)	12 (4.7)	33 (8.0)	26 (7.1)	27 (6.7)	
6-10 days	15 (3.5)	24 (5.9)	11 (3.3)	17 (6.6)	22 (5.3)	19 (5.2)	26 (6.4)	
11-30 days	39 (9.0)	43 (10.6)	45 (13.3)	22 (8.6)	49 (11.8)	36 (9.9)	38 (9.4)	
31-100 days	68 (15.7)	48 (11.8)	43 (12.7)	38 (14.8)	55 (13.3)	51 (14.0)	50 (12.4)	
101-365 days	88 (20.4)	93 (22.9)	76 (22.5)	66 (25.7)	86 (20.7)	81 (22.3)	101 (25.0)	
>365 days	184 (42.6)	164 (40.4)	133 (39.3)	98 (38.1)	160 (38.6)	141 (38.7)	155 (38.4)	

[a] Chi-squared tests used to calculate *P* values.

[b] Data collected on the same weekday but during 2 separate weeks. For example, Monday[b] indicates data collected on Mondays from 2 different weeks.

[c] Data collected on a weekday on a single week. For example, Thursday[b] indicates data collected on a Thursday from a single week of data collection.

[d] Sixth and 7th grade combined in education variable, and Asian and Pacific Islander combined in race to conduct chi-squared tests.

[e] n=14 respondents never smoked cannabis (ie, n=14 had only ever used an alternative method of administration such as vaping or edible).

Discussion and Future Directions

The utilization of social media–based recruitment for conducting cannabis epidemiology research has only begun to scratch the surface of its potential. Our experiences in conducting this research have revealed several insights about this method and the resulting data. First, it would appear prudent to operate under the assumption that Facebook advertising algorithms require a sufficient amount of time to learn about a target population to effectively disseminate advertisements. Thus, we have found it useful to wait at least 48 to 72 hours before evaluating the effectiveness of a Facebook advertising campaign. Second, Facebook advertising methods that recruit for Web-based surveys on cannabis use can quickly and inexpensively generate samples of current, regular cannabis users across the United States. Given the speed of data collection, we have also found that this sampling method is useful for timely clarification of hypotheses, via the expeditious conduct of follow-up surveys. Overall, we encourage the conceptualization of Web survey dissemination via Facebook advertising as a new type of nonprobability-based targeted sampling method. Cannabis researchers can include Facebook as part of their armamentarium of sampling methods. It can be considered as an alternative or adjunct to traditional survey methods, perhaps most valuable when trying to study cannabis behaviors not captured by traditional methods. In addition, in light of the difficulties in developing data infrastructures for evaluating the effects of cannabis legalization [73] and literature suggesting that changes to the legal status of cannabis may disproportionately impact the behavior of regular cannabis users [74], Facebook recruitment methods appear to be valuable for collecting policy-relevant data.

Web surveying via Facebook is only one of many emerging epidemiological methods for studying cannabis use. A growing body of literature has demonstrated that digital trace data [44] (ie, records of naturalistic digital behavior and communication) can be used to study cannabis use. For example, Twitter has been used to study new forms of cannabis administration such as dabbing [75,76], vaping [77], and edibles [78,79]; perceptions, attitudes, and normalization of cannabis use [80-82]; and unique communities of cannabis user subgroups and network structures of cannabis dispensaries [83,84]. Data from other Web-based platforms such as Reddit [85], Instagram [86], YouTube [87], and search engines such as Google [88] or Bing [89] are all other potential sources of digital trace data that have been used to study patterns of cannabis use.

Given the sensitive nature of substance use data, the concept of IP address tracking warrants discussion. When designing a survey, researchers must consider the balance between the need to prevent multiple responses and the need for anonymity to obtain valid responses. Without being able to determine who has already taken a survey, researchers run the risk of having single individuals complete a survey multiple times—especially if monetary incentives are used to encourage survey participation [90]. However, failure to provide respondents with anonymity may result in fewer responses or invalid data. In our surveys to date, we have used the prevent ballot box stuffing feature of Qualtrics as a middle ground. This feature uses cookies rather than IP addresses to prevent multiple responses, which allows us to maintain participant anonymity. Even so, the promise of anonymity may not convince individuals involved in illicit behavior to participate [91]. Finally, note that the use of IP addresses for preventing multiple responses can inadvertently prevent individuals who share the same IP address (eg, college dorm, shared house) from completing the survey.

The landscape of social media will continue to evolve in the coming years, and specific social media platforms such as Facebook may not remain as popular as they are today. However, all indications are that digital social networking and mass communication platforms are here to stay and will continue to grow as advertising tools. In 2017, businesses spent approximately 13.5 billion dollars using social media marketing to sell their products and services [92]. Thus, it seems likely that providing an advertising service within digital social networking platforms will remain a staple means of generating revenue for the companies that create such platforms. In addition, as the machine learning techniques used to disseminate digital marketing advertisements continue to become more sophisticated and effective, researchers can leverage such advances to even more efficiently reach and collect data on clinical subpopulations of interest.

There remain many unanswered questions related to this sampling method. Here, we mention just 3, each of which could provide exciting additional research opportunities. First, can these methods be used effectively to study the use of other substances? It is likely that the degree of stigma surrounding a particular substance of interest will affect the utility of the methodology presented here. For example, there are nationally distributed magazines and well-recognized political activist groups devoted specifically to the promotion of cannabis use and cannabis culture. These aspects of cannabis culture are available as specific targeted interests on the Facebook Business advertising platform. In contrast, to our knowledge, there are no regularly published national magazines or political activist groups that promote crystal methamphetamine use and culture. Thus, the targeting strategies for more stigmatized drugs such as crystal methamphetamine may need to be different and will likely require the use of monetary incentives. One suggestion to overcome this issue is to conduct an in-person pilot focus group with regular users of the substance of interest and collect digital or self-report data concerning their Web-based behavior. The resulting data could then be used to generate a more effective Facebook advertising campaign.

Second, will participants provide identifying contact information in a survey to facilitate longitudinal survey follow-up (ie, repeated sampling from the same individual over time)? Previous work by Harris et al has demonstrated that survey respondents recruited via Facebook are willing to provide contact information to facilitate follow-up for longitudinal studies. However, their study aimed to understand patterns of contraception use among young Australian women [30]. It remains to be seen whether something similar can be accomplished with a US-based sample of heavy cannabis users.

Finally, is it possible to use alternative advertising strategies (eg, not using cannabis-related targeted interests) to obtain samples of less experienced or less frequent cannabis users? Preliminary data currently being collected by our team suggest that this can be readily accomplished using different targeting strategies, but that data collection may be somewhat slower and more expensive.

Sampling issues are a primary concern of any epidemiological investigation because they dictate the conclusions that can be drawn from the data [63,64]. Additional methodological evaluations of social media–based sampling will fill essential gaps in our knowledge of how to use the data effectively. Presently, we believe that social media–based Web surveys have tremendous utility for members of the research community and will continue to facilitate our understanding of the evolving nature of cannabis use behaviors.

Acknowledgments

The authors would like to thank Maria Stoianova for reviewing the paper and providing helpful edits and suggestions. Funding for this work was provided by NIDA: R01-DA032243, R01-DA015186, T32-DA037202, P30-DA029926. The National Institutes of Health had no role in study design; in the collection, analysis and interpretation of data; in the writing of the report; or in the decision to submit the paper for publication.

Conflicts of Interest

None declared.

References

1. Lee BK. Epidemiologic research and Web 2.0--the user-driven Web. Epidemiology 2010 Nov;21(6):760-763. [doi: 10.1097/EDE.0b013e3181f5a75f] [Medline: 20924229]
2. Morabia A. A History of Epidemiologic Methods and Concepts. Switzerland: Birkhäuser Basel; 2004.
3. Cooper SL. Random sampling by telephone: an improved method. J Mark Res 1964 Nov 21;1(4):45-48. [doi: 10.2307/3150377]
4. Rothman KJ, Cann CI, Walker AM. Epidemiology and the internet. Epidemiology 1997 Mar;8(2):123-125. [Medline: 9229201]
5. van Gelder MM, Bretveld RW, Roeleveld N. Web-based questionnaires: the future in epidemiology? Am J Epidemiol 2010 Dec 01;172(11):1292-1298 [FREE Full text] [doi: 10.1093/aje/kwq291] [Medline: 20880962]
6. Miller PG, Sønderlund AL. Using the internet to research hidden populations of illicit drug users: a review. Addiction 2010 Sep;105(9):1557-1567. [doi: 10.1111/j.1360-0443.2010.02992.x] [Medline: 20626378]
7. Coomber R. Dangerous drug adulteration-an international survey of drug dealers using the internet and the world wide web (WWW). Int J Drug Policy 1997;8:71-81 [FREE Full text]
8. Barratt M, Lenton S. Beyond recruitment? Participatory online research with people who use drugs. Int J Internet Res Ethics 2010;3(1):69-86.
9. Statista. 2017. Number of internet users worldwide from 2005 to 2016 URL: https://www.statista.com/statistics/273018/number-of-internet-users-worldwide/ [accessed 2017-05-04] [WebCite Cache ID 6urFlI2K5]
10. eMarketer. Statista. 2017. Number of smartphone users worldwide from 2014 to 2020 (in billions) URL: https://www.statista.com/statistics/330695/number-of-smartphone-users-worldwide/ [accessed 2017-11-10] [WebCite Cache ID 6urG24hqL]
11. eMarketer. Statista. 2017. Number of social network users worldwide from 2010 to 2021 (in billions) URL: https://www.statista.com/statistics/278414/number-of-worldwide-social-network-users/ [accessed 2017-11-10] [WebCite Cache ID 6urGEnPRx]
12. Statista. 2017. Dossier on Facebook URL: https://www.statista.com/study/9711/facebook-statista-dossier/ [accessed 2017-11-10] [WebCite Cache ID 6usAkcbAM]
13. Statista. 2017. Dossier: Social media usage worldwide URL: https://www.statista.com/study/12393/social-networks-statista-dossier/ [accessed 2018-02-18] [WebCite Cache ID 6xKlIwhu0]
14. Statista. 2017. Daily time spent on social networking by internet users worldwide from 2012 to 2017 (in minutes) URL: https://www.statista.com/statistics/433871/daily-social-media-usage-worldwide/ [accessed 2018-02-18] [WebCite Cache ID 6xKkp5pz9]
15. Stewart J. Nytimes. New York, NY: The New York Times; 2016 May 05. Facebook Has 50 minutes of Your Time Each Day It Wants More URL: https://www.nytimes.com/2016/05/06/business/facebook-bends-the-rules-of-audience-engagement-to-its-advantage.html [accessed 2017-08-10] [WebCite Cache ID 6urGrgsZN]

16. Topolovec-Vranic J, Natarajan K. The use of social media in recruitment for medical research studies: a scoping review. J Med Internet Res 2016 Nov 7;18(11):e286 [FREE Full text] [doi: 10.2196/jmir.5698] [Medline: 27821383]
17. Thornton L, Batterham PJ, Fassnacht DB, Kay-Lambkin F, Calear AL, Hunt S. Recruiting for health, medical or psychosocial research using Facebook: Systematic review. Internet Interv 2016 May;4:72-81. [doi: 10.1016/j.invent.2016.02.001]
18. Frandsen M, Thow M, Ferguson SG. The effectiveness of social media (Facebook) compared with more traditional advertising methods for recruiting eligible participants to health research studies: a randomized, controlled clinical trial. JMIR Res Protoc 2016;5(3):e161 [FREE Full text] [doi: 10.2196/resprot.5747] [Medline: 27511829]
19. Eysenbach G. Improving the quality of Web surveys: the Checklist for Reporting Results of Internet E-Surveys (CHERRIES). J Med Internet Res 2004 Sep 29;6(3):e34 [FREE Full text] [doi: 10.2196/jmir.6.3.e34] [Medline: 15471760]
20. Fan WM, Yan Z. Factors affecting response rates of the web survey: a systematic review. Comput Human Behav 2010 Mar;26(2):132-139. [doi: 10.1016/j.chb.2009.10.015]
21. Sánchez-Fernández J, Muñoz-Leiva F, Montoro-Ríos FJ. Improving retention rate and response quality in Web-based surveys. Comput Human Behav 2012 Mar;28(2):507-514. [doi: 10.1016/j.chb.2011.10.023]
22. Eysenbach G, Wyatt J. Using the Internet for surveys and health research. J Med Internet Res 2002 Nov;4(2):E13 [FREE Full text] [doi: 10.2196/jmir.4.2.e13] [Medline: 12554560]
23. Manfreda KL, Berzelak J, Vehovar V, Bosnjak M, Haas I. Web surveys versus other survey Modes: a meta-analysis comparing response rates. Int J Mark Res 2008 Jan;50(1):79-104. [doi: 10.1177/147078530805000107]
24. Shih TH, Fan XT. Comparing response rates from web and mail surveys: a meta-analysis. Field Methods 2008 Mar 03;20(3):249-271. [doi: 10.1177/1525822X08317085]
25. Whitaker C, Stevelink S, Fear N. The use of Facebook in recruiting participants for health research purposes: a systematic review. J Med Internet Res 2017 Aug 28;19(8):e290 [FREE Full text] [doi: 10.2196/jmir.7071] [Medline: 28851679]
26. Yuan P, Bare MG, Johnson MO, Saberi P. Using online social media for recruitment of human immunodeficiency virus-positive participants: a cross-sectional survey. J Med Internet Res 2014;16(5):e117 [FREE Full text] [doi: 10.2196/jmir.3229] [Medline: 24784982]
27. Tustin JL, Crowcroft NS, Gesink D, Johnson I, Keelan J, Lachapelle B. Facebook recruitment of vaccine-hesitant Canadian parents: cross-sectional study. JMIR Public Health Surveill 2017 Jul 24;3(3):e47 [FREE Full text] [doi: 10.2196/publichealth.6870] [Medline: 28739557]
28. Nelson EJ, Hughes J, Oakes JM, Pankow JS, Kulasingam SL. Estimation of geographic variation in human papillomavirus vaccine uptake in men and women: an online survey using facebook recruitment. J Med Internet Res 2014;16(9):e198 [FREE Full text] [doi: 10.2196/jmir.3506] [Medline: 25231937]
29. Kapp JM, Peters C, Oliver DP. Research recruitment using Facebook advertising: big potential, big challenges. J Cancer Educ 2013 Mar;28(1):134-137. [doi: 10.1007/s13187-012-0443-z] [Medline: 23292877]
30. Harris ML, Loxton D, Wigginton B, Lucke JC. Recruiting online: lessons from a longitudinal survey of contraception and pregnancy intentions of young Australian women. Am J Epidemiol 2015 May 15;181(10):737-746. [doi: 10.1093/aje/kwv006] [Medline: 25883155]
31. Batterham PJ. Recruitment of mental health survey participants using Internet advertising: content, characteristics and cost effectiveness. Int J Methods Psychiatr Res 2014 Jun;23(2):184-191. [doi: 10.1002/mpr.1421] [Medline: 24615785]
32. Crosier BS, Brian RM, Ben-Zeev D. Using Facebook to reach people who experience auditory hallucinations. J Med Internet Res 2016;18(6):e160 [FREE Full text] [doi: 10.2196/jmir.5420] [Medline: 27302017]
33. Chu JL, Snider CE. Use of a social networking web site for recruiting Canadian youth for medical research. J Adolesc Health 2013 Jun;52(6):792-794. [doi: 10.1016/j.jadohealth.2012.12.002] [Medline: 23352727]
34. Lord S, Brevard J, Budman S. Connecting to young adults: an online social network survey of beliefs and attitudes associated with prescription opioid misuse among college students. Subst Use Misuse 2011;46(1):66-76 [FREE Full text] [doi: 10.3109/10826084.2011.521371] [Medline: 21190407]
35. Jones SC, Robinson L, Barrie L, Francis K, Lee JK. Association between young Australian's drinking behaviours and their interactions with alcohol brands on Facebook: results of an online survey. Alcohol Alcohol 2016 Jul;51(4):474-480. [doi: 10.1093/alcalc/agv113] [Medline: 26487157]
36. Ramo DE, Hall SM, Prochaska JJ. Reliability and validity of self-reported smoking in an anonymous online survey with young adults. Health Psychol 2011 Nov;30(6):693-701 [FREE Full text] [doi: 10.1037/a0023443] [Medline: 21574709]
37. Ramo DE, Young-Wolff KC, Prochaska JJ. Prevalence and correlates of electronic-cigarette use in young adults: findings from three studies over five years. Addict Behav 2015 Feb;41:142-147. [doi: 10.1016/j.addbeh.2014.10.019] [Medline: 25452058]
38. Lenton S, Frank VA, Barratt MJ, Dahl HV, Potter GR. Attitudes of cannabis growers to regulation of cannabis cultivation under a non-prohibition cannabis model. Int J Drug Policy 2015 Mar;26(3):257-266. [doi: 10.1016/j.drugpo.2014.08.002] [Medline: 25181993]
39. Barratt MJ, Lenton S. Representativeness of online purposive sampling with Australian cannabis cultivators. Int J Drug Policy 2015 Mar;26(3):323-326. [doi: 10.1016/j.drugpo.2014.10.007] [Medline: 25433986]
40. Barratt MJ, Bouchard M, Decorte T, Asmussen Frank V, Hakkarainen P, Lenton S, et al. Understanding global patterns of domestic cannabis cultivation. Drugs Alcohol Today 2012 Nov 30;12(4):213-221. [doi: 10.1108/17459261211286627]

41. Ramo DE, Prochaska JJ. Prevalence and co-use of marijuana among young adult cigarette smokers: an anonymous online national survey. Addict Sci Clin Pract 2012 Apr 19;7:5 [FREE Full text] [doi: 10.1186/1940-0640-7-5] [Medline: 23186143]
42. Ramo DE, Delucchi KL, Hall SM, Liu H, Prochaska JJ. Marijuana and tobacco co-use in young adults: patterns and thoughts about use. J Stud Alcohol Drugs 2013 Mar;74(2):301-310 [FREE Full text] [Medline: 23384378]
43. Eysenbach G. Infodemiology and infoveillance: framework for an emerging set of public health informatics methods to analyze search, communication and publication behavior on the Internet. J Med Internet Res 2009;11(1):e11 [FREE Full text] [doi: 10.2196/jmir.1157] [Medline: 19329408]
44. Salathé M, Bengtsson L, Bodnar TJ, Brewer DD, Brownstein JS, Buckee C, et al. Digital epidemiology. PLoS Comput Biol 2012;8(7):e1002616 [FREE Full text] [doi: 10.1371/journal.pcbi.1002616] [Medline: 22844241]
45. Ekman A, Litton JE. New times, new needs; e-epidemiology. Eur J Epidemiol 2007;22(5):285-292. [doi: 10.1007/s10654-007-9119-0] [Medline: 17505896]
46. ProCon. Medicalmarijuana.procon. 29 Legal Medical Marijuana States and DC URL: https://medicalmarijuana.procon.org/view.resource.php?resourceID=000881 [accessed 2017-11-10] [WebCite Cache ID 6urH53vHe]
47. Carlini BH, Garrett SB, Harwick RM. Beyond joints and brownies: marijuana concentrates in the legal landscape of WA State. Int J Drug Policy 2017 Apr;42:26-29. [doi: 10.1016/j.drugpo.2017.01.004] [Medline: 28171805]
48. Barrus DG, Capogrossi KL, Cates SC, Gourdet CK, Peiper NC, Novak SP, et al. Tasty THC: promises and challenges of cannabis edibles. Methods Rep RTI Press 2016 Nov;2016 [FREE Full text] [Medline: 28127591]
49. Budney AJ, Sargent JD, Lee DC. Vaping cannabis (marijuana): parallel concerns to e-cigs? Addiction 2015 Nov;110(11):1699-1704. [doi: 10.1111/add.13036] [Medline: 26264448]
50. Loflin M, Earleywine M. A new method of cannabis ingestion: the dangers of dabs? Addict Behav 2014 Oct;39(10):1430-1433. [doi: 10.1016/j.addbeh.2014.05.013] [Medline: 24930049]
51. Alshaikh F, Ramzan F, Rawaf S, Majeed A. Social network sites as a mode to collect health data: a systematic review. J Med Internet Res 2014;16(7):e171 [FREE Full text] [doi: 10.2196/jmir.3050] [Medline: 25048247]
52. Kosinski M, Matz SC, Gosling SD, Popov V, Stillwell D. Facebook as a research tool for the social sciences: opportunities, challenges, ethical considerations, and practical guidelines. Am Psychol 2015 Sep;70(6):543-556. [doi: 10.1037/a0039210] [Medline: 26348336]
53. Umbach PD. Web surveys: best practices. In: Crisp G, editor. New Directions for Institutional Research. Hoboken, NJ: Wiley; 2004:23-38.
54. Wang C, Raina R, Fong D, Zhou D, Han J, Badros G. Learning Relevance from a Heterogeneous Social Network and Its Application in Online Targeting. 2011 Presented at: ACM Special Interest Group on Information Retrieval (SIGIR); July 24-28, 2011; Beijing.
55. Xinran H, Pan J, Jin O, Xu T, Liu B, Xu T. Practical Lessons from Predicting Clicks on Ads at Facebook. 2014 Presented at: International Workshop on Data Mining for Online Advertising (ADKDD); August 24-27, 2014; New York, NY. [doi: 10.1145/2648584.2648589]
56. Hazelwood K, Bird S, Brooks D, Chintala S, Diril U, Dzhulgakov D. Applied Machine Learning at Facebook: A Datacenter Infrastructure Perspective. 2018 Presented at: International Symposium on High-Performance Computer Architecture (HPCA); February 24-28, 2018; Vienna, Austria. [doi: 10.1109/HPCA.2018.00059]
57. Borodovsky JT, Budney AJ. Legal cannabis laws, home cultivation, and use of edible cannabis products: a growing relationship? Int J Drug Policy 2017 Nov 02;50:102-110. [doi: 10.1016/j.drugpo.2017.09.014] [Medline: 29102847]
58. Facebook Inc. Facebook. 2018. About the delivery system: pacing URL: https://www.facebook.com/business/help/1754368491258883?helpref=related&ref=related# [accessed 2018-01-21]
59. Borodovsky JT, Crosier BS, Lee DC, Sargent JD, Budney AJ. Smoking, vaping, eating: is legalization impacting the way people use cannabis? Int J Drug Policy 2016 Oct;36:141-147 [FREE Full text] [doi: 10.1016/j.drugpo.2016.02.022] [Medline: 26992484]
60. Borodovsky JT, Lee DC, Crosier BS, Gabrielli JL, Sargent JD, Budney AJ. U.S. cannabis legalization and use of vaping and edible products among youth. Drug Alcohol Depend 2017 Aug 01;177:299-306. [doi: 10.1016/j.drugalcdep.2017.02.017] [Medline: 28662974]
61. Lee DC, Crosier BS, Borodovsky JT, Sargent JD, Budney AJ. Online survey characterizing vaporizer use among cannabis users. Drug Alcohol Depend 2016 Feb 01;159:227-233 [FREE Full text] [doi: 10.1016/j.drugalcdep.2015.12.020] [Medline: 26774946]
62. Sloboda Z. Defining and measuring drug abusing behaviors. In: Sloboda Z, editor. Epidemiology of Drug Abuse. Boston, MA: Springer US; 2005:3-14.
63. Kakinami L, Conner KR. Sampling strategies for addiction research. In: Miller PG, Strang J, Miller PM, editors. Addiction Research Methods. United Kingdom: Blackwell Publishing Ltd; 2010:27-42.
64. Taylor C, Griffiths P. Sampling issues in drug epidemiology. In: Sloboda Z, editor. Epidemiology of Drug Abuse. Boston, MA: Springer US; 2005:79-98.
65. Fitch C, Rhodes T, Hope V, Stimson G, Renton A. The role of rapid assessment methods in drug use epidemiology. In: The Science of Drug Abuse Epidemiology. Vienna, Austria: United Nations Office on Drugs and Crime; 2003.

66. Dupont HB, Kaplan CD, Braam RV, Verbraeck HT, de Vries NK. The application of the rapid assessment and response methodology for cannabis prevention research among youth in the Netherlands. Int J Drug Policy 2015 Aug;26(8):731-738 [FREE Full text] [doi: 10.1016/j.drugpo.2014.11.003] [Medline: 25481612]
67. Center for Behavioral Health Statistics and Quality. National Survey on Drug Use and Health 2015. Research Triangle Park, NC: RTI International and United States Department of Health and Human Services. Substance Abuse and Mental Health Services Administration; 2016.
68. Borodovsky J, Lee DC, Crosier BS, Budney AJ. Cannabis use data collected via social media: a methodological comparison. Drug Alcohol Depend 2017 Feb;171:e24. [doi: 10.1016/j.drugalcdep.2016.08.081]
69. Barratt MJ, Ferris JA, Zahnow R, Palamar JJ, Maier LJ, Winstock AR. Moving on from representativeness: testing the utility of the Global Drug Survey. Subst Abuse 2017;11:1178221817716391 [FREE Full text] [doi: 10.1177/1178221817716391] [Medline: 28924351]
70. U.S. Census Bureau. Census. Annual Estimates of the Resident Population April 1, 2010 to July 1, 2016 URL: https://www.census.gov/data/datasets/2017/demo/popest/state-total.html [accessed 2018-04-06] [WebCite Cache ID 6yTpmody1]
71. Brick JM, Kalton G. Handling missing data in survey research. Stat Methods Med Res 1996 Sep;5(3):215-238. [doi: 10.1177/096228029600500302] [Medline: 8931194]
72. Ramo DE, Liu H, Prochaska JJ. Reliability and validity of young adults' anonymous online reports of marijuana use and thoughts about use. Psychol Addict Behav 2012 Dec;26(4):801-811 [FREE Full text] [doi: 10.1037/a0026201] [Medline: 22082344]
73. Kilmer B, Pacula RL. Building the data infrastructure to evaluate cannabis legalization. Addiction 2017 Jul;112(7):1140-1141. [doi: 10.1111/add.13824] [Medline: 28477353]
74. Pacula RL, Lundberg R. Why changes in price matter when thinking about marijuana policy: a review of the literature on the elasticity of demand. Public Health Rev 2014;35(2):1-18 [FREE Full text] [Medline: 25642015]
75. Daniulaityte R, Nahhas RW, Wijeratne S, Carlson RG, Lamy FR, Martins SS, et al. “Time for dabs”: analyzing Twitter data on marijuana concentrates across the U.S. Drug Alcohol Depend 2015 Oct 1;155:307-311. [doi: 10.1016/j.drugalcdep.2015.07.1199] [Medline: 26338481]
76. Cavazos-Rehg PA, Sowles SJ, Krauss MJ, Agbonavbare V, Grucza R, Bierut L. A content analysis of tweets about high-potency marijuana. Drug Alcohol Depend 2016 Sep 01;166:100-108. [doi: 10.1016/j.drugalcdep.2016.06.034] [Medline: 27402550]
77. Sowles SJ, Krauss MJ, Connolly S, Cavazos-Rehg PA. A content analysis of vaping advertisements on Twitter, November 2014. Prev Chronic Dis 2016 Sep 29;13:E139 [FREE Full text] [doi: 10.5888/pcd13.160274] [Medline: 27685432]
78. Cavazos-Rehg PA, Zewdie K, Krauss MJ, Sowles SJ. “No high like a brownie high”: a content analysis of edible marijuana tweets. Am J Health Promot 2018;32(4):880-886. [doi: 10.1177/0890117116686574] [Medline: 29214836]
79. Lamy FR, Daniulaityte R, Sheth A, Nahhas RW, Martins SS, Boyer EW, et al. “Those edibles hit hard”: exploration of Twitter data on cannabis edibles in the U.S. Drug Alcohol Depend 2016 Jul 1;164:64-70. [doi: 10.1016/j.drugalcdep.2016.04.029] [Medline: 27185160]
80. Thompson L, Rivara FP, Whitehill JM. Prevalence of marijuana-related traffic on Twitter, 2012-2013: a content analysis. Cyberpsychol Behav Soc Netw 2015 Jun;18(6):311-319 [FREE Full text] [doi: 10.1089/cyber.2014.0620] [Medline: 26075917]
81. Cavazos-Rehg P, Krauss M, Grucza R, Bierut L. Characterizing the followers and tweets of a marijuana-focused Twitter handle. J Med Internet Res 2014;16(6):e157 [FREE Full text] [doi: 10.2196/jmir.3247] [Medline: 24974893]
82. Cavazos-Rehg PA, Krauss M, Fisher SL, Salyer P, Grucza RA, Bierut LJ. Twitter chatter about marijuana. J Adolesc Health 2015 Feb;56(2):139-145. [doi: 10.1016/j.jadohealth.2014.10.270] [Medline: 25620299]
83. Peiper NC, Baumgartner PM, Chew RF, Hsieh YP, Bieler GS, Bobashev GV, et al. Patterns of Twitter behavior among networks of cannabis dispensaries in California. J Med Internet Res 2017 Jul 04;19(7):e236 [FREE Full text] [doi: 10.2196/jmir.7137] [Medline: 28676471]
84. Baumgartner P, Peiper N. Utilizing big data and Twitter to discover emergent online communities of cannabis users. Subst Abuse 2017;11:1178221817711425 [FREE Full text] [doi: 10.1177/1178221817711425] [Medline: 28615950]
85. Sowles SJ, Krauss MJ, Gebremedhn L, Cavazos-Rehg PA. “I feel like I've hit the bottom and have no idea what to do”: supportive social networking on Reddit for individuals with a desire to quit cannabis use. Subst Abus 2017;38(4):477-482. [doi: 10.1080/08897077.2017.1354956] [Medline: 28704167]
86. Cavazos-Rehg PA, Krauss MJ, Sowles SJ, Bierut LJ. Marijuana-related posts on Instagram. Prev Sci 2016 Aug;17(6):710-720. [doi: 10.1007/s11121-016-0669-9] [Medline: 27262456]
87. Krauss MJ, Sowles SJ, Mylvaganam S, Zewdie K, Bierut LJ, Cavazos-Rehg PA. Displays of dabbing marijuana extracts on YouTube. Drug Alcohol Depend 2015 Oct 1;155:45-51. [doi: 10.1016/j.drugalcdep.2015.08.020] [Medline: 26347408]
88. Zhang Z, Zheng X, Zeng DD, Leischow SJ. Tracking dabbing using search query surveillance: a case study in the United States. J Med Internet Res 2016 Sep 16;18(9):e252 [FREE Full text] [doi: 10.2196/jmir.5802] [Medline: 27637361]
89. Yom-Tov E, Lev-Ran S. Adverse reactions associated with cannabis consumption as evident from search engine queries. JMIR Public Health Surveill 2017 Oct 26;3(4):e77 [FREE Full text] [doi: 10.2196/publichealth.8391] [Medline: 29074469]

90. Bowen AM, Daniel CM, Williams ML, Baird GL. Identifying multiple submissions in Internet research: preserving data integrity. AIDS Behav 2008 Feb 1;12(6):964-973. [doi: 10.1007/s10461-007-9352-2]
91. Barratt MJ, Potter GR, Wouters M, Wilkins C, Werse B, Perälä J, et al. Lessons from conducting trans-national Internet-mediated participatory research with hidden populations of cannabis cultivators. Int J Drug Policy 2015 Mar;26(3):238-249. [doi: 10.1016/j.drugpo.2014.12.004] [Medline: 25576247]
92. Statista. 2017. Dossier Social Media Marketing in the U.S URL: https://www.statista.com/study/15449/social-media-marketing-in-the-us-statista-dossier/ [accessed 2017-11-10] [WebCite Cache ID 6usAwhlIY]

Abbreviations

IP: Internet protocol
LCL: legal cannabis laws
NSDUH: National Survey on Drug Use and Health

Edited by G Eysenbach; submitted 13.11.17; peer-reviewed by R Daniulaityte, T Kowatsch, M Barratt; comments to author 05.12.17; revised version received 26.01.18; accepted 24.02.18; published 02.05.18.

Please cite as:
Borodovsky JT, Marsch LA, Budney AJ
Studying Cannabis Use Behaviors With Facebook and Web Surveys: Methods and Insights
JMIR Public Health Surveill 2018;4(2):e48
URL: http://publichealth.jmir.org/2018/2/e48/
doi:10.2196/publichealth.9408
PMID:29720366

Publisher:
JMIR Publications
130 Queens Quay East.
Toronto, ON, M5A 3Y5
Phone: (+1) 416-583-2040
Email: support@jmir.org