JMIR Public Health and Surveillance

Impact Factor (2022): 8.5
Volume 1 (2015), Issue 2 ISSN 2369-2960 Editor in Chief: Travis Sanchez, PhD, MPH

Contents

Original Papers

Protocol

Original Paper

The US National Tuberculosis Surveillance System: A Descriptive Assessment of the Completeness and Consistency of Data Reported from 2008 to 2012

Rachel S Yelk Woodruff[1*], MPH; Robert H Pratt[1*], BSIE; Lori R Armstrong[1*], PhD

Centers for Disease Control and Prevention, National Center for HIV/AIDS, Viral Hepatitis, STD, and TB Prevention, Division of Tuberculosis Elimination, Atlanta, GA, United States

[*]all authors contributed equally

Corresponding Author:
Rachel S Yelk Woodruff, MPH
Centers for Disease Control and Prevention
National Center for HIV/AIDS, Viral Hepatitis, STD, and TB Prevention
Division of Tuberculosis Elimination
1600 Clifton Road NE, Mailstop E10
Atlanta, GA, 30333
United States
Phone: 1 404 639 6018
Fax: 1 404 639 8958
Email: zex5@cdc.gov

Abstract

Background: In 2009, the Tuberculosis (TB) Information Management System transitioned into the National TB Surveillance System to allow use of 4 different types of electronic reporting schemes: state-built, commercial, and 2 schemes developed by the Centers for Disease Control and Prevention. Simultaneously, the reporting form was revised to include additional data fields.

Objective: Describe data completeness for the years 2008-2012 and determine the impact of surveillance changes.

Methods: Data were categorized into subgroups and assessed for completeness (eg, the percentage of patients dead at diagnosis who had a date of death reported) and consistency (eg, the percentage of patients alive at diagnosis who erroneously had a date of death reported). Reporting jurisdictions were grouped to examine differences by reporting scheme.

Results: Each year less than 1% of reported cases had missing information for country of origin, race, or ethnicity. Patients reported as dead at diagnosis had death date (a new data field) missing for 3.6% in 2009 and 4.4% in 2012. From 2010 to 2012, 313 cases (1%) reported as alive at diagnosis had a death date and all of these were reported through state-built or commercial systems. The completeness of reporting for guardian country of birth for pediatric patients (a new data field) ranged from 84% in 2009 to 88.2% in 2011.

Conclusions: Despite major changes, completeness has remained high for most data elements in TB surveillance. However, some data fields introduced in 2009 remain incomplete; continued training is needed to improve national TB surveillance data.

(JMIR Public Health Surveill 2015;1(2):e15) doi:10.2196/publichealth.4991

KEYWORDS

public health surveillance; disease notification; information systems; data cleaning; quality assurance

Introduction

Tuberculosis (TB) incidence (or case notification) is used globally for monitoring trends, planning, and evaluating public health programs [1,2]. In the United States, national incidence reporting began in 1953, with documented cases and operational data from each reporting jurisdiction submitted in aggregate [3]. By 1985, all jurisdictions were reporting individual cases using a standardized form, the Report of Verified Case of Tuberculosis (RVCT) [4]. In 1993, the RVCT was expanded to include additional risk factors and laboratory information, and TB surveillance data began to be entered and transmitted to the Centers for Disease Control and Prevention (CDC) through a single software system [5].

The US National Tuberculosis Surveillance System (NTSS) underwent major revisions in 2009 [6]. RVCT was expanded

to include 11 new data fields, and 25 of 38 existing fields were modified. Concurrently, state and local reporting areas transitioned from reporting TB case data through the Tuberculosis Information Management System (TIMS), a stand-alone, modem-based system developed at the CDC, to their choice of 4 reporting schemes: (1) the National Electronic Disease Surveillance System (NEDSS)-base system, a CDC-developed infrastructure; (2) the electronic RVCT (eRVCT), also developed by the CDC; (3) state-developed custom software systems; or (4) commercial software developed by private companies. All reporting schemes were required to conform to specific Public Health Information Network and NEDSS data standards [7,8].

The transition from a single reporting scheme to a choice of different types of schemes allowed state and local TB programs more control over the structure of their surveillance systems and gave them responsibility for their own data validation [9]. Prior to 2009, surveillance data came to the CDC via TIMS, which had a built-in data validation system for alerting logic errors to help ensure accurate data entry and reporting. These validation standards were retired with TIMS in 2010, although the CDC-developed eRVCT and NEDSS-base system retained validation rules similar to those in TIMS. Validation rules for state-developed and commercial schemes vary by jurisdiction. Furthermore, routine maintenance, updates, changes, and enhancements of state-developed and commercial reporting schemes are now at the expense of state and local TB programs; information technology (IT) expertise is necessary at the state and local level to maintain and update these types of systems [9]. Modifications of state and commercial reporting schemes, such as changes in RVCT data fields, have to be done at the level of the individual reporting jurisdiction; therefore, modifications to NTSS are more complicated than they were prior to 2009, when the CDC was able to update a single system and provide all reporting jurisdictions with updated software that incorporated the revisions.

The objectives of this report are to describe the completeness and consistency of TB case data reported to the CDC from 2008 to 2012, to determine the extent to which the 2009 changes in RVCT and reporting schemes affected the data, and to find ways to improve data quality. Although the surveillance report and the reporting schemes described here are specific to TB, the analytical methods and results may be useful to managers of other public health programs who are contemplating similar changes in surveillance systems or reporting schemes.

Methods

Data sources

NTSS receives TB surveillance data electronically from the 50 states and the District of Columbia [6]. The reporting officials in TB programs collect laboratory and clinical TB data from a variety of sources and store them in electronic reporting systems. From 1998 to 2009, those officials submitted TB surveillance data through TIMS by using file-transfer protocol and controlled-access Internet and modem transfer [10]. Starting in 2009, TB surveillance data have been transmitted using Public Health Information Network Messaging Service software in HL7 messaging format.

The CDC provides preliminary TB surveillance datasets weekly for reporting program officials to verify reported data. The CDC creates final TB surveillance datasets annually for reporting, research, and publications. Since 2009, TB data reported to the CDC have been subjected to a data-cleaning routine before a finalized dataset is created. The data cleaning routine is applied to selected data fields using a hierarchical strategy as determined by CDC staff (eg, a dependent field, such as the year of previous TB episode, is deleted if the independent field, such as history of previous TB, is not present) that creates a dataset that has fewer inconsistencies but not necessarily more accuracy. Our analysis included only clean, finalized annual datasets.

Analysis

We examined responses from NTSS data elements from 2008 to 2012 (the most recent year of data at the time of analysis) and new elements from 2009 to 2012. Although NTSS includes data from 1993 to 2012, the purpose of this study was to examine how the changes in data elements and reporting schemes affected the data; therefore, the study period begins the year before the changes occurred. New data elements from Alaska, California, Connecticut, Illinois, Missouri, Mississippi, North Carolina, North Dakota, New York City, and Ohio were not included for 2009 because these jurisdictions used TIMS that year and the new elements were not supported. In addition, we excluded California and Vermont from analyses that included HIV test results for 2008-2012 because HIV reporting practices were different for these jurisdictions.

Reporting jurisdictions were categorized according to the type of reporting scheme (TIMS, commercial, eRVCT, NEDSS-base, or state-developed) used in 2009 and 2010-2012. Because of the changes in both reporting schemes and RVCT in 2009, data from that year were examined separately from latter years' data.

Data were categorized into subgroups and data elements associated with subgroups were assessed for completeness (eg, the percentage of patients dead at diagnosis who had a date of death reported) and consistency (eg, the percentage of patients alive at diagnosis who erroneously had a date of death reported). The results are presented for a subset of data elements that are clinically or demographically important or exhibited inconsistency or incompleteness in reporting. Furthermore, for each TB case we selected key data elements from 3 different categories: risk factors, clinical aspects of TB disease, and molecular aspects of TB disease.

Results

From 2008 to 2012, 56,040 cases were reported to NTSS [6]. Each year, fewer than 1% of reported cases had missing or unknown information for origin of birth (nativity; 59/56,040), or race/ethnicity (197/56,040). One data element that demonstrated inconsistency in completeness was correctional facility status (residence in correctional facility at time of diagnosis), for which 6.5% of cases (746/11,520) had unknown or missing information in 2009, compared with approximately 1% or less of cases (265/44,529) in other years (Table 1). When

correctional facility status was examined by reporting system (Table 2), information was missing for 17.1% (729/4266) of the cases reported by jurisdictions using TIMS in 2009, while the other reporting systems had less than 1% of cases (17/6871) missing for this element. Among cases reported as residents in correctional facilities at the time of diagnosis, information on the type of correctional facility was missing for 9% (10/110) of cases reported through state-developed reporting systems in 2009 and 2010-2012 (25/267), compared to less than 3% (17/1386) through TIMS, commercial, NEDSS-based, and eRVCT reporting systems for those same years (Tables 2 and 3).

Table 1. Completeness of trend data elements reported to the National Tuberculosis Surveillance System, United States, 2008-2012.

		2008		2009		2010		2011		2012	
		N	%	N	%	N	%	N	%	N	%
Total reported TB cases		12,904	100.0	11,520	100.0	11,163	100.0	10,517	100.0	9945	100.0
Resident in correctional facility											
	Yes	499	3.9	465	4.0	489	4.4	423	4.0	386	3.9
	No	12,386	96.0	10,309	89.5	10,536	94.4	10,036	95.4	9509	95.6
	Unknown/ Missing	19	0.1	746	6.5	138	1.2	58	0.6	50	0.5
Type of correctional facility indicated[a]		492	98.6	451	97.0	471	96.3	411	97.2	382	99.0
History of TB		572	4.4	492	4.3	510	4.6	511	4.9	481	4.8
Year of TB reported[b]		564	98.6	455	92.5	493	96.7	497	97.3	462	96.0
Year of TB missing[b]		8	1.4	37	7.5	17	3.3	14	2.7	19	4.0
Initial DST done[c]		9604	98.4	8725	98.2	8316	98.4	7966	98.5	7315	96.3
Isoniazid results[c]		9385	97.7	8684	99.5	8279	99.6	7923	99.5	7258	99.2
Rifampin results[c]		9377	97.6	8678	99.5	8279	99.6	7919	99.4	7260	99.3

[a]Among patients who were residents of correctional facilities at the time of diagnosis.

[b]Among cases that reported history of previous TB.

[c]Drug susceptibility test. Among patients who had positive culture; includes resistant and susceptible test results.

Table 2. Completeness and consistency of data elements reported to the National Tuberculosis Surveillance System by type of reporting system, United States, 2009.

		TIMS[b]		Commercial		State developed		NEDSS-base[c]		eRVCT[d]	
		N	%	N	%	N	%	N	%	N	%
Total reported TB cases		4266	100.0	470	100.0	2815	100.0	2852	100.0	1117	100.0
Resident in correctional facility											
	No	3438	80.6	463	98.5	2689	95.5	2654	93.1	1065	95.3
	Unknown/ Missing	729	17.1	0	0.0	16	0.6	1	0.0	0	0.0
	Yes	99	2.3	7	1.5	110	3.9	197	6.9	52	4.7
Type of correctional facility indicated[a]		92	92.9	7	100.0	100	90.9	197	100.0	51	98.1
History of TB		222	100.0	20	100.0	130	100.0	89	100.0	31	100.0
Year of TB reported		195	87.8	18	90.0	124	95.4	88	98.9	30	96.8
Year of TB missing		27	12.2	2	10.0	6	4.6	1	1.1	1	3.2
Alive at time of TB diagnosis		4182	100.0	466	100.0	2747	100.0	2787	100.0	1094	100.0
Date of death indicated		N/A	N/A	8	1.7	40	1.5	0	0.0	0	0.0
Patient < 15 years of age		N/A	N/A	32	100.0	153	100.0	214	100.0	44	100.0
Guardian country of birth		N/A	N/A	28	87.5	119	77.8	189	88.3	36	81.8
Patient 15 years of age or older		N/A	N/A	438	100.0	2662	100.0	2638	100.0	1073	100.0
Guardian country of birth		N/A	N/A	1	0.2	63	2.4	8	0.3	2	0.2

[a]Among patients who were residents at correctional facilities at the time of diagnosis.

[b]Tuberculosis Information Management System.

[c]National Electronic Disease Surveillance System.

[d]Electronic Report of Verified Case of Tuberculosis.

Table 3. Completeness and consistency of data elements reported to the National Tuberculosis Surveillance System by type of reporting system, United States, 2010-2012.

	Commercial		State developed		NEDSS-base[a]		eRVCT[b]	
	N	%	N	%	N	%	N	%
Total reported TB cases	12,685	100.0	8551	100.0	7646	100.0	2743	100.0
Resident in correctional institute	312	100.0	267	100.0	489	100.0	230	100.0
Type of correctional facility indicated	303	97.1	242	90.6	489	100.0	230	100.0
History of TB	689	100.0	421	100.0	286	100.0	106	100.0
Year of TB	665	96.5	408	96.9	276	96.5	103	97.2
Alive at time of TB diagnosis	12,397	100.0	8330	100.0	7468	100.0	2680	100.0
Date of death indicated	162	1.3	151	1.8	0	0.0	0	0.0
Patient < 15 years of age	597	100.0	484	100.0	504	100.0	119	100.0
Guardian country of birth	525	87.9	374	77.3	485	96.2	105	88.2
Patient 15 years of age or older	12,088	100.0	8067	100.0	7142	100.0	2624	100.0
Guardian country of birth	38	0.3	317	3.9	19	0.3	3	0.1

[a]National Electronic Disease Surveillance System.

[b]Electronic Report of Verified Case of Tuberculosis.

In 2009, 7.5% of cases (37/492) with a previous history of TB reported were missing the year previous TB disease occurred, compared to 1.4% (8/572) in 2008 (Table 1). No previous year of TB disease was reported for cases that did not have a history of previous TB disease indicated. Among cases reported in 2009 with a previous history of TB disease indicated, the highest percentage of missing years of previous TB disease was with TIMS at 12.2% (27/222; Table 2), compared to 10% or less (10/270) of cases with a previous history of TB that were missing years of previous TB disease reported through the other systems (Table 2). For 2010-2012, the year of previous TB disease was missing for 3-4% of cases (50/1502) for which previous TB disease history was indicated across all reporting system types (Table 3).

Of the 426 culture-positive cases reported in 2008 that did not have initial drug susceptibility testing (4.2% of all culture-positive cases, 426/10,024, including those with unknown or missing initial drug susceptibility test results), 1 case was reported as susceptible to isoniazid and 1 case was reported as susceptible to rifampin. From 2009 to 2012, no culture-positive cases without initial drug susceptibility test reported "done" had isoniazid or rifampin results reported. For sputum culture and sputum smear results reported as negative or positive, over 99% of cases (31,098/31,410) each year had a sputum smear or sputum culture collection date reported (Table 4). No sputum culture or sputum smear collection dates were reported for cases that did not have an associated sputum culture or sputum smear test done.

Table 4. Completeness of new data elements reported to the National Tuberculosis Surveillance System, United States, 2009-2012.

	2009		2010		2011		2012	
	N	%	N	%	N	%	N	%
Dead at time of TB diagnosis	160	100.0	252	100.0	245	100.0	221	100.0
Date of death indicated	153	95.6	244	96.8	235	95.9	213	96.4
Sputum culture positive or negative	5743	100.0	9018	100.0	8599	100.0	8050	100.0
Sputum collection date indicated	5704	99.3	8952	99.3	8502	98.9	7940	98.6
Sputum smear positive or negative	5788	100.0	9162	100.0	8709	100.0	8217	100.0
Sputum smear date indicated	5743	99.2	9139	99.7	8674	99.6	8141	99.1
Patient < 15 years of age	443	100.0	637	100.0	578	100.0	489	100.0
Guardian country of birth	372	84.0	555	87.1	510	88.2	424	86.7
Lived outside US > 2 months	120	27.1	161	25.3	124	21.5	139	28.4
Country where lived indicated[a]	116	96.7	151	93.8	119	96.0	135	97.1

[a] Among pediatric patients who lived outside the country for 2 months.

For cases reported as dead at TB diagnosis, 4.4% (7/160) were missing date of death in 2009, the first year date of death information was collected, and 4.6% (8/221) were missing it in 2012 (Table 4). In 2009, 48 of 7094 TB cases (0.70%) were reported as alive at diagnosis and had a date of death indicated (Table 5). A majority of these (83%, 40/48; Table 2) were reported through state-developed systems. From 2010 to 2012, 313 of 30,875 TB cases (1%) were reported as alive at diagnosis and had a date of death indicated (Table 5); all were reported through state-developed or commercial reporting systems (Table 3).

Table 5. Consistency between new data elements reported to the National Tuberculosis Surveillance System, United States, 2009–2012.

	2009		2010		2011		2012	
	N	%	N	%	N	%	N	%
Alive at time of TB diagnosis	7094	100.0	10,903	100.0	10,261	100.0	9711	100.0
Date of death indicated	48	0.7	111	1.0	116	1.1	86	0.9
Sputum culture not done/unknown/missing	1511	100.0	2100	100.0	1851	100.0	1748	100.0
Sputum collection date indicated	5	0.3	0	0.0	0	0.0	0	0.0
Sputum smear not done/unknown/missing	1466	100.0	1998	100.0	1799	100.0	1703	100.0
Sputum smear date indicated	1	0.1	0	0.0	0	0.0	0	0.0
Patient 15 years of age or older	6811	100.0	10,526	100.0	9939	100.0	9456	100.0
Guardian country of birth	74	1.1	102	1.0	152	1.5	123	1.3
Lived outside US > 2 months	101	1.5	178	1.7	183	1.8	157	1.7
Country where lived indicated[a]	92	91.1	168	94.4	179	97.8	155	98.7

[a]Among pediatric patients who lived outside the country for more than 2 months.

Country of birth of primary guardian, whether the patient lived outside the United States for more than 2 months and if so in what countries, are new data elements requested for pediatric patients (<15 years of age). Completeness ranged from 84% (372/443) in 2009 to 88.2% (510/578) in 2011 for the guardian country of birth for pediatric TB cases and from 93.8% (151/161) in 2010 to 97.1% (135/139) in 2012 for the country where the pediatric patient lived for more than 2 months (Table 4). Among nonpediatric cases (15 years of age and older), 1-2% (451/36,732) each year indicated a country of birth for the primary guardian. In 2009 and 2010-2012, completeness in reporting for guardian country of birth for pediatric TB patients was highest for those reported through NEDSS-base software systems (88.3%, 189/214, and 96.2%, 485/504, respectively; Tables 2 and 3). Nonpediatric cases with primary guardian information were predominantly reported through state-developed software systems in 2009 (Table 2) and 2010-2012 (Table 3).

Discussion

Principal Findings

Considering the extent of changes the US TB Surveillance System underwent in 2009, TB surveillance data have maintained a high level of completeness, with most data elements showing the same levels of completeness after 2009. New data elements, for which collection and reporting began in 2009 for most reporting jurisdictions, have varied completeness but show an overall improvement from 2009 to 2012. Some new data elements are taking longer to reach a high percentage of completeness at the state and local levels, or are less complete or less concordant in 2012 than they were in 2009. For example, patients who were dead at the time of TB diagnosis should have had a corresponding date of death recorded (the date-of-death data element was introduced in 2009). However, some jurisdictions reported a date of death for patients who were alive at diagnosis, which occurred more frequently in 2012 than in 2009 (Table 5). If a patient is alive at TB diagnosis and dies during therapy, there is no corresponding date of death field; therefore, some reporting jurisdictions may be recording the date of patient death in the field for death date of patients who were dead at the time of TB diagnosis. Among cases reported in 2009 that were alive at diagnosis and had a date of death recorded, 58% (28/48) had a date of death that matched the date therapy was stopped (data not shown), indicating that the date of death field was used to record the date of death during therapy. Completeness may also have been affected by lack of information or inability to find information in patient records, misinterpretation of data element definitions, or use of a paper reporting form that does not match the electronic reporting data entry form [2]. For some jurisdictions, electronic reporting systems may not have been revised to accommodate reporting of certain data elements; therefore, those elements cannot be reported electronically. Ongoing training of local staff to account for turnover and changes in duties may improve completeness of reporting [2].

The data cleaning routine does not take into consideration all possible data errors. Information requested specifically for all TB patients less than 15 years of age was sometimes reported for cases 15 years of age or older (Tables 2, 3, and 5), and the date of death may have been indicated for patients who were alive at diagnosis (Tables 2, 3, and 5); these discrepancies are not corrected as part of data cleaning. Therefore, care is warranted when working with NTSS data for reporting or research purposes. Proper subsetting is needed to prevent inclusion of patients who should not be included in a specific subset for analysis, such as patients alive at diagnosis when analyzing date of death, as these exclusions are not built into the dataset and omitting them could result in erroneous results.

Differences in completeness of data reported through the different electronic systems may be due to system configuration or reporting practices within the jurisdictions. The high percentage of missing correctional facility information reported in 2009 (Table 1) was due to data transmission problems experienced by a single reporting jurisdiction. The information for residence in a correctional facility existed in TIMS but was not transferred from TIMS to the jurisdiction's new reporting

system. Furthermore, commercial and state-developed reporting systems are responsible for their own validation, which could account for some higher percentages of missing or inaccurate data. TB case surveillance data do not allow for assessment of systems or reporting practices at the state and local level, so it was not possible to distinguish between factors related to systems or reporting practices in this analysis.

In 2009 there was an unexpected and significant decline in the numbers of TB cases reported to NTSS compared to previous years [11]. Changes to electronic reporting systems were not deemed to be a causal factor. Rather, we concluded that the decline in TB cases was a result of decreased TB diagnoses in the United States. Therefore, we did not consider the unexpected decline in TB cases in 2009 to be a factor in our study.

Limitations

This study has several limitations. Limited resources prevented us from conducting a validation study at the local level to compare patient data from medical charts to the data reported to NTSS. This would have been especially valuable to assess data elements that exhibited inconsistency. The data-cleaning routine replaced some validation rules that existed in TIMS but may not have improved the quality of data reported to the CDC. For example, from 2009 to 2012, 2 cases reported as not having initial susceptibility testing done were also reported as susceptible to both isoniazid and rifampin (data not shown), indicating that initial drug susceptibility testing may actually have been done. Because the cases were reported as not undergoing susceptibility testing, the susceptibility results were deleted for these cases during data cleaning and therefore are not reflected in the clean, finalized dataset. Isoniazid and rifampin are important drugs for treating TB and resistance to both defines multidrug-resistant TB. If susceptibility testing was indeed done for isoniazid and rifampin, then drug susceptibility testing should be reported as “done” on RVCT.

Conclusion

Several ongoing efforts have been implemented to improve the quality of surveillance reporting. The CDC initiated a series of trainings in 2010 with the goal of familiarizing state and local reporting jurisdictions with the updated RVCT and reporting requirements [12]. Additionally, in 2011, the CDC conducted a series of trainings on quality assurance of TB data [13]. The trainings culminated in a published manual that is available to reporting jurisdictions and others interested in attaining high-quality surveillance data [14]. A collection of reports showing various aspects of TB data reported to the CDC is available through NTSS to authorized state and local TB program staff. Information provided through NTSS reports includes the numbers of missing and unknown values associated with reported data elements, the frequency of reporting for select elements, when data were last transmitted to the CDC, and a list of elements with no information ever reported for a particular reporting area. State and local TB program staff can use these reports to identify and correct gaps in reported data or to report data errors to the CDC. The National Tuberculosis Indicators Project (NTIP) can also be used to verify and check TB surveillance data reported to the CDC [13]. Reporting jurisdictions can compare their records with NTIP data and use the NTIP to identify discrepancies. The RVCT has an accompanying manual that provides comprehensive reporting guidance for each data element [15]. Furthermore, the RVCT workgroup, composed of CDC and state and local TB program staff, actively pursues clarification and provides guidance on improving RVCT reporting. As state and local TB control programs are often challenged with declining resources and staff turnover, the CDC should periodically provide updated quality assurance and RVCT training webinars and materials to ensure that TB control program staff remain aware of data problem areas and new and existing quality assurance tools and techniques. These efforts, as well as ongoing discussions regarding data quality assurance, will improve the completeness and accuracy of TB surveillance data.

State and local communicable disease surveillance systems vary from disease-specific systems to systems used for reporting an array of diseases and conditions [9]. However, from 2007 to 2010, interoperability and integration of state and local public health disease surveillance systems increased substantially [9]. As public health programs begin to utilize current advances in electronic reporting and embrace new national guidelines related to health information exchange and meaningful use, more electronic surveillance systems will be modified to increase capacity and meet national standards [9,16]. The results of the NTSS transition from a single, stand-alone surveillance system to a variety of different reporting schemes illustrate that major modifications of disease surveillance systems can be done without substantial impact on the completeness of surveillance data.

Conflicts of Interest

None declared.

References

1. World Health Organization. Global Tuberculosis Report. Geneva, Switzerland: World Health Organization; 2013. URL: http://apps.who.int/iris/bitstream/10665/91355/1/9789241564656_eng.pdf [accessed 2015-09-16] [WebCite Cache ID 6basTNaiX]
2. Sprinson JE, Lawton ES, Porco TC, Flood JM, Westenhouse JL. Assessing the validity of tuberculosis surveillance data in California. BMC Public Health 2006;6:217 [FREE Full text] [doi: 10.1186/1471-2458-6-217] [Medline: 16930492]
3. Public Health Service (US). Reported tuberculosis morbidity and other data. In: Reported Tuberculosis Data 1953-69. Washington, DC: US Department of Health, Education, and Welfare; 1953.

4. Centers for Disease Control and Prevention (US). 1985 Tuberculosis statistics, states and cities. In: Tuberculosis Statistics in the United States, 1985, 1985-1986, 1987, 1988. Atlanta, GA: US Department of Health and Human Services; 1986:1-89.
5. Centers for Disease Control and Prevention (US). Reported Tuberculosis in the United States. Atlanta, GA: US Department of Health and Human Services; 1993.
6. Centers for Disease Control and Prevention (US). Reported Tuberculosis in the United States, 2012. Atlanta, GA: US Department of Health and Human Services; 2012. URL: http://www.cdc.gov/tb/statistics/reports/2012/ [accessed 2015-09-16] [WebCite Cache ID 6batJhFHd]
7. Centers for Disease Control and Prevention (US). Public Health Information Network. URL: http://www.cdc.gov/phin/about/index.html [accessed 2015-07-28] [WebCite Cache ID 6aMgdFfL5]
8. Centers for Disease Control and Prevention (US). National Notifiable Diseases Surveillance System (NNDSS). Atlanta, GA: Centers for Disease Control and Prevention URL: http://wwwn.cdc.gov/nndss/nedss.html [accessed 2015-07-28] [WebCite Cache ID 6aMqfHNf4]
9. Centers for Disease Control and Prevention (US). State electronic disease surveillance systems—United States, 2007 and 2010. MMWR Morb Mortal Wkly Rep 2011 Oct 21;60(41):1421-1423 [FREE Full text] [Medline: 22012115]
10. National Public Health Service and Biosurveillance Registry for Human Health (US). Tuberculosis Information Management System (TIMS). Atlanta, GA: US Department of Health and Human Services; 2011.
11. Winston CA, Navin TR, Becerra JE, Chen MP, Armstrong LR, Jeffries C, et al. Unexpected decline in tuberculosis cases coincident with economic recession—United States, 2009. BMC Public Health 2011;11:846 [FREE Full text] [doi: 10.1186/1471-2458-11-846] [Medline: 22059421]
12. Magee E, Tryon C, Forbes A, Heath B, Manangan L. The National Tuberculosis Surveillance System training program to ensure accuracy of tuberculosis data. J Public Health Manag Pract 2011;17(5):427-430. [doi: 10.1097/PHH.0b013e31820f8e43] [Medline: 21788780]
13. Manangan LP, Tryon C, Magee E, Miramontes R. Innovative quality-assurance strategies for tuberculosis surveillance in the United States. Tuberc Res Treat 2012;2012:481230 [FREE Full text] [doi: 10.1155/2012/481230] [Medline: 22685648]
14. Centers for Disease Control and Prevention (US). Quality assurance for tuberculosis surveillance data: A guide and toolkit. 2009. URL: ftp://ftp.cdc.gov/pub/Software/TIMS/2009%20RVCT%20Documentation/RVCT%20Training%20Materials/Quality%20Assurance%20Materials/QA%20Manual/ [accessed 2015-03-13] [WebCite Cache ID 6bcRhinAR]
15. Centers for Disease Control and Prevention (US). CDC tuberculosis surveillance data training: Report of Verified Case of Tuberculosis (RVCT) instruction manual. Atlanta, GA: Centers for Disease Control and Prevention (US) URL: http://www.cdc.gov/tb/programs/rvct/instructionmanual.pdf [accessed 2015-07-28] [WebCite Cache ID 6aMshVik5]
16. Shapiro JS, Mostashari F, Hripcsak G, Soulakis N, Kuperman G. Using health information exchange to improve public health. Am J Public Health 2011 Apr;101(4):616-623. [doi: 10.2105/AJPH.2008.158980] [Medline: 21330598]

Abbreviations

CDC: Centers for Disease Control and Prevention
eRVCT: electronic Report of Verified Case of Tuberculosis
NEDSS: National Electronic Diseases Surveillance System
NTIP: National Tuberculosis Indicators Project
NTSS: National Tuberculosis Surveillance System
RVCT: Report of Verified Case of Tuberculosis
TB: tuberculosis
TIMS: Tuberculosis Information Management System

Edited by T Sanchez; submitted 30.07.15; peer-reviewed by N Baruch, N Nordin; comments to author 22.08.15; revised version received 07.09.15; accepted 11.09.15; published 15.10.15.

Please cite as:
Yelk Woodruff RS, Pratt RH, Armstrong LR
The US National Tuberculosis Surveillance System: A Descriptive Assessment of the Completeness and Consistency of Data Reported from 2008 to 2012
JMIR Public Health Surveill 2015;1(2):e15
URL: http://publichealth.jmir.org/2015/2/e15/
doi:10.2196/publichealth.4991
PMID:27227133

Original Paper

Implementation of a Multimodal Mobile System for Point-of-Sale Surveillance: Lessons Learned From Case Studies in Washington, DC, and New York City

Jennifer Cantrell[1,2], MPA, DrPH; Ollie Ganz[1,3], MSPH; Vinu Ilakkuvan[1,3], MSPH; Michael Tacelosky[4]; Jennifer Kreslake[2], MPH, PhD; Joyce Moon-Howard[5], MPH, DrPH; Angela Aidala[6], PhD; Donna Vallone[1,5], MPH, PhD; Andrew Anesetti-Rothermel[7], PhD; Thomas R Kirchner[2,5,8], PhD

[1]Truth Initiative, Evaluation Science and Research, Washington, DC, United States

[2]Department of Health, Behavior, and Society, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, United States

[3]Department of Prevention and Community Health, George Washington University Milken Institute School of Public Health, Washington, DC, United States

[4]Survos LLC, Washington, DC, United States

[5]College of Global Public Health, New York University, New York, NY, United States

[6]Department of Sociomedical Sciences, Columbia University Mailman School of Public Health, New York, NY, United States

[7]Steven A Schroeder Institute for Tobacco Research and Policy Studies, Truth Initiative, Washington, DC, United States

[8]Department of Oncology, Lombardi Comprehensive Cancer Center, Georgetown University Medical Center, Washington, NY, United States

Corresponding Author:
Jennifer Cantrell, MPA, DrPH
Truth Initiative
Evaluation Science and Research
900 G Street, NW, Fourth Floor
Washington, DC, 20001
United States
Phone: 1 202 436 2118
Fax: 1 202 454 5798
Email: jcantrell@truthinitiative.org

Abstract

Background: In tobacco control and other fields, point-of-sale surveillance of the retail environment is critical for understanding industry marketing of products and informing public health practice. Innovations in mobile technology can improve existing, paper-based surveillance methods, yet few studies describe in detail how to operationalize the use of technology in public health surveillance.

Objective: The aims of this paper are to share implementation strategies and lessons learned from 2 tobacco, point-of-sale surveillance projects to inform and prepare public health researchers and practitioners to implement new mobile technologies in retail point-of-sale surveillance systems.

Methods: From 2011 to 2013, 2 point-of-sale surveillance pilot projects were conducted in Washington, DC, and New York, New York, to capture information about the tobacco retail environment and test the feasibility of a multimodal mobile data collection system, which included capabilities for audio or video recording data, electronic photographs, electronic location data, and a centralized back-end server and dashboard. We established a preimplementation field testing process for both projects, which involved a series of rapid and iterative tests to inform decisions and establish protocols around key components of the project.

Results: Important components of field testing included choosing a mobile phone that met project criteria, establishing an efficient workflow and accessible user interfaces for each component of the system, training and providing technical support to fieldworkers, and developing processes to integrate data from multiple sources into back-end systems that can be utilized in real-time.

Conclusions: A well-planned implementation process is critical for successful use and performance of multimodal mobile surveillance systems. Guidelines for implementation include (1) the need to establish and allow time for an iterative testing

framework for resolving technical and logistical challenges; (2) developing a streamlined workflow and user-friendly interfaces for data collection; (3) allowing for ongoing communication, feedback, and technology-related skill-building among all staff; and (4) supporting infrastructure for back-end data systems. Although mobile technologies are evolving rapidly, lessons learned from these case studies are essential for ensuring that the many benefits of new mobile systems for rapid point-of-sale surveillance are fully realized.

(JMIR Public Health Surveill 2015;1(2):e20) doi:10.2196/publichealth.4191

KEYWORDS

mobile technology; public health surveillance; tobacco; point-of-sale; implementation; tobacco industry advertising; marketing

Introduction

Public health surveillance is necessary for collecting health data and for planning and evaluating programs and policies [1]. In tobacco control and other fields, point-of-sale surveillance of the retail environment is critical for understanding industry marketing of products and informing public health practice [2]. Innovations in mobile technologies provide opportunities to improve existing surveillance methods. Recent studies comparing mobile versus paper-based data collection for surveillance have identified several benefits of mobile technologies, including reduced data loss [3], real-time quality control [3], rapid use of data [3,4], and reduced costs [4]. Mobile phones have the added benefit of being less conspicuous than paper surveys [4]—an important consideration in retail surveillance.

Previous retail assessments of tobacco marketing and related public health topics such as alcohol marketing and food availability have traditionally involved paper surveys completed by trained data collectors within a sample of representative stores [5-9]. More recent studies mention other modes of data collection for tobacco point-of-sale assessment [10-13], but few report their methodology with any detail [6] and none have described the challenges and potential solutions for implementing multimodal mobile systems that incorporate data from different sources. A detailed and well-planned implementation process is critical to ensuring surveillance systems gather up-to-date, accurate information quickly and efficiently, which is especially critical for low-resource environments.

This paper aims to address this gap in the literature by sharing implementation lessons from 2 point-of-sale surveillance projects, with the objective of preparing public health practitioners for the process of implementing multimodal systems using mobile technologies for point-of-sale marketing assessments. First, we describe the background of each project and the mobile systems and tools used. We then describe the preimplementation field testing process, providing examples of challenges and solutions to issues confronted during testing. We summarize by providing lessons learned for implementing mobile technologies for point-of-sale surveillance.

Project Descriptions

From August 2011 to May 2013, our research group conducted 2 point-of-sale surveillance projects in Washington, DC, and New York City. The goals of these projects were (1) to capture point-of-sale marketing information to educate communities about tobacco marketing in their neighborhoods and inform policy and (2) to assess the feasibility of multimodal mobile data collection tools for point-of-sale surveillance.

Washington, DC, Point-of-Sale Surveillance Project

The Washington, DC, project was funded by the District of Columbia (DC) Department of Health and Truth Initiative (known as the "American Legacy Foundation" at the time of this project) and ran from August 2011 to March 2012. This project was a traditional research study, in which researchers hired and trained paid professional fieldworkers to collect data on tobacco advertising in all outlets licensed to sell tobacco (n=1061) in DC, a midsized urban city of approximately 650,000 people, with a large African-American population. More details on the study can be found elsewhere [14-17].

New York City Point-of-Sale Surveillance Project

The New York City (NYC) study was funded by Truth Initiative and was conducted from February to May 2013. This study was a community-based participatory research project and involved a comprehensive survey of all outlets within selected census tracts in Central Harlem, a predominantly low- to moderate-income, African-American community with a total of approximately 156 outlets that serve a local population of 165,000 people over a 2.1-mi^2 area [18-20]. The study included the formation of an academic-community partnership between researchers from Truth Initiative, Columbia University's Mailman School of Public Health, and community representatives with experience in tobacco control work in Harlem. Local students recruited through a nonprofit Boys & Girls Club were paid a small stipend for conducting data collection. More details on this study can be found elsewhere [21].

Both projects were designed to include some or all of the following components:

1. *Audio or voice recording data:* Responses to survey questions (by selecting or typing a response or by voice recording) on tobacco advertising on the exterior and/or interior of outlets;
2. *Electronic photographs:* Images of tobacco advertisements in outlets;
3. *Electronic location data:* Information about the outlet, such as the store name, address, and/or GPS coordinates.

Surveillance System

Prior to the commencement of the DC and NYC projects, the authors worked with an information technology service provider

to develop a flexible multimodal surveillance system to collect data through digital, or nonpaper-based, methods. The system was based on a server-client model in which data could be entered via a mobile device or computer (ie, acting as the "client") and communicate directly and in real-time with the remote server. Data were transmitted via a wireless local area network or via a mobile phone network.

Software

Software allowed for data collection via mobile technologies utilizing text messaging, email, GPS technologies, and phone-based interactive voice response (IVR), a technology that allows a computer to interact with humans through the use of voice and touchtone input via keypad. The IVR software was written in PHP (a Web-based programming language) and integrated with a telephone provider (Voxeo), communicating via the VoiceXML protocol, which is designed specifically to enable IVR and remote computer systems to exchange data. In addition, a test version of a custom mobile application functional for both iOS and Android operating systems was developed prior to the NYC project. This app was not available for the DC project.

Server and Dashboard

All data from the field were automatically sent to and stored in a secure server. The server provided authorized users with access to a Web interface that contained tools for visualizing and managing data.

Methods

Data Collection Modes

For each project, we chose tools from the multimodal system and software based on project needs, software availability at the time, fieldworker expertise, and the environment in which the project was implemented. For the DC study, primary data collection modes included an IVR-programmed survey taken using any mobile phone and photographs taken with a mobile phone with a built-in camera and emailing capabilities. For NYC, tools included an IVR-programmed survey accessed via mobile phone and the aforementioned mobile app on a smart phone used to take photographs.

A preimplementation field period involved conducting a series of iterative tests to inform decisions and establish protocols around the components of the project, described below. The DC field testing period lasted 7 weeks and included field testing, survey instrument development, and reliability testing. Applying lessons learned in DC, our field testing period for NYC was approximately 2 weeks.

Mobile Phone Selection

Field tests involved photographic or survey data collection at a limited number of stores using a specific mobile phone, with fieldworkers reporting on problems encountered in the field with the device, using that feedback to inform the next series of tests, and repeating the process until final mobile phone criteria were established.

Workflow and User Interface

For both projects, the workflow involved unobtrusively taking photographs and collecting survey data on tobacco advertising for each store exterior and interior. The user interface in this study referred to the fieldworker interaction with the IVR survey, phone cameras, phone email, and mobile app. Different workflows were tested along with varying aspects of the user interface until the most efficient process was established.

For fieldworker training and technical support, we allotted time for initial training on the specifics of tobacco advertising data collection as well as practicing with the technologies, testing the workflow and interface in the field, and establishing a feedback loop between fieldworkers and researchers.

For the back-end data system, researchers and fieldworkers tested different strategies for linking survey, photo, and location data collected from various modes to an individual store, with different solutions developed for each project. As part of the linking and geocoding process, we obtained lists of all businesses licensed to sell tobacco in the area from the city governments, which included business names and addresses. We geocoded addresses for each study utilizing the Master Repository Geocoder from the DC Office of the Chief Technology Officer [22], a free tool available for addresses within the District. For NYC, we utilized ArcGIS Online geocoding services [23], which were free at the time. The research team also worked with the technologist to develop a custom-made, secure online website that allowed for real-time posting and monitoring of data.

To come to decisions on key issues, the full interdisciplinary team—including researchers, project managers, technologists, and field staff—met formally 1-2 times per week to discuss relevant components of the project, identify the specific challenges and decisions to be addressed, and work through each issue utilizing a group consensus-based approach. Relevant materials were provided before or during the meeting. During the user interface and website development phases, the technologist team provided versions of the survey via IVR or updates to the website to be tested by team members and subsequently discussed during meetings to resolve major issues and decide on next steps. Additional discussions were held on a Web forum established during the pilot process to resolve field testing issues (described below) and over email. Smaller or more urgent decisions were often made on a daily basis using these electronic channels or daily debriefs with fieldworkers. Preliminary lessons learned were decided upon by the team and based on the first DC project; they were then further refined after the NYC project for the purposes of this paper.

Results

Below we discuss the main challenges and solutions that resulted from the field testing process, including examples to illustrate how technical and other challenges were resolved for project implementation.

Choosing a Mobile Phone

The choice of a mobile phone will facilitate or hinder usability, visibility, and quality of data collection, thus criteria for phone

selection should be established based on the type of data to be collected and following a series of field tests. For both projects, most mobile phones with cellular service could be used for IVR survey data collection. However, the collection of multiple high-quality photographs and linking photos with survey and store identification data required further criteria for mobile phone selection. Criteria established included the following:

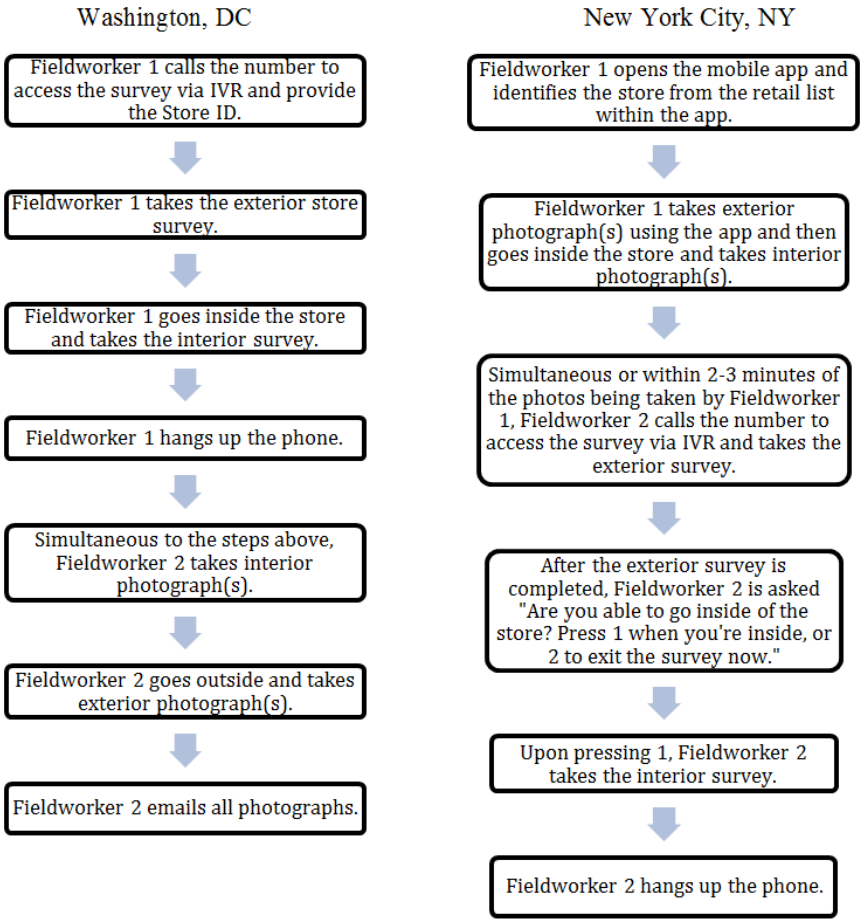
1. *Wireless technologies:* A minimum of 3G cellular connectivity was required. For the NYC project, phones also required Wi-Fi connectivity so fieldworkers could quickly upload photos to the server from the mobile app via a wireless connection as it was often too slow to upload photos over the cellular network.
2. *Operating system:* Either an Android or iOS operating system was required for compatibility with apps that were in development or in use during the projects.
3. *Camera:* A rear-facing camera with a minimum of 5-8 megapixels was required. A higher number of megapixels indicates higher quality photographs. Camera features should allow for the ability to disable the flash and shutter noise, which is essential for being unobtrusive while taking photos in the retail environment. The camera must also allow for geotagging of photos when GPS is turned on.
4. *Display:* The display technology of the phone had to allow for high resolution visibility outdoors and in direct sunlight, preferably with a Super AMOLED touchscreen.
5. *Battery life:* Phone battery life for talk time had to extend through the length of at least one field shift, which was usually 4-6 hours long.
6. *Global positioning system (GPS):* The phone required GPS-enabled receiving capability.
7. *Other features:* Phone had to have a headphone jack for listening to the IVR survey and reducing background noise.

Based on the aforementioned criteria, we chose the Samsung Focus in 2011 for the DC project and the Samsung Galaxy S II 4G in 2012 for the NYC project.

Workflow and Front-End User Interface

The field testing process demonstrated that utilizing fieldworker pairs to survey stores was needed for safety purposes and to optimize workflow. Streamlining the field data collection necessitated that the fieldworkers share tasks at each store and required allowing for time and flexibility during data collection to respond to delays or unexpected events. Figure 1 outlines the workflow used in DC and NYC. Based on lessons learned in DC, we added questions to the survey for NYC that allowed fieldworkers time to pause the IVR survey when needed to move around naturally (ie, from the exterior of the store to the interior).

Figure 1. Final optimized workflow for DC and NYC.



User Interface: IVR Survey

In developing a streamlined user interface, we established 3 key guiding principles: simplicity, accuracy, and flexibility. Fieldworkers tested different stages of the user interface in the field and evaluated the interaction in light of these principles. The most challenging interface was the user's interaction with the IVR for the purpose of taking the survey. Audio surveys taken via IVR take longer than pencil-and-paper surveys and make complicated survey questions difficult. IVR questions were tested in training sessions and in the field to identify improvements for shortening each question item's stem and response categories and improving overall survey flow. Table 1 provides examples of the application of the 3 principles to resolve challenges with the IVR survey.

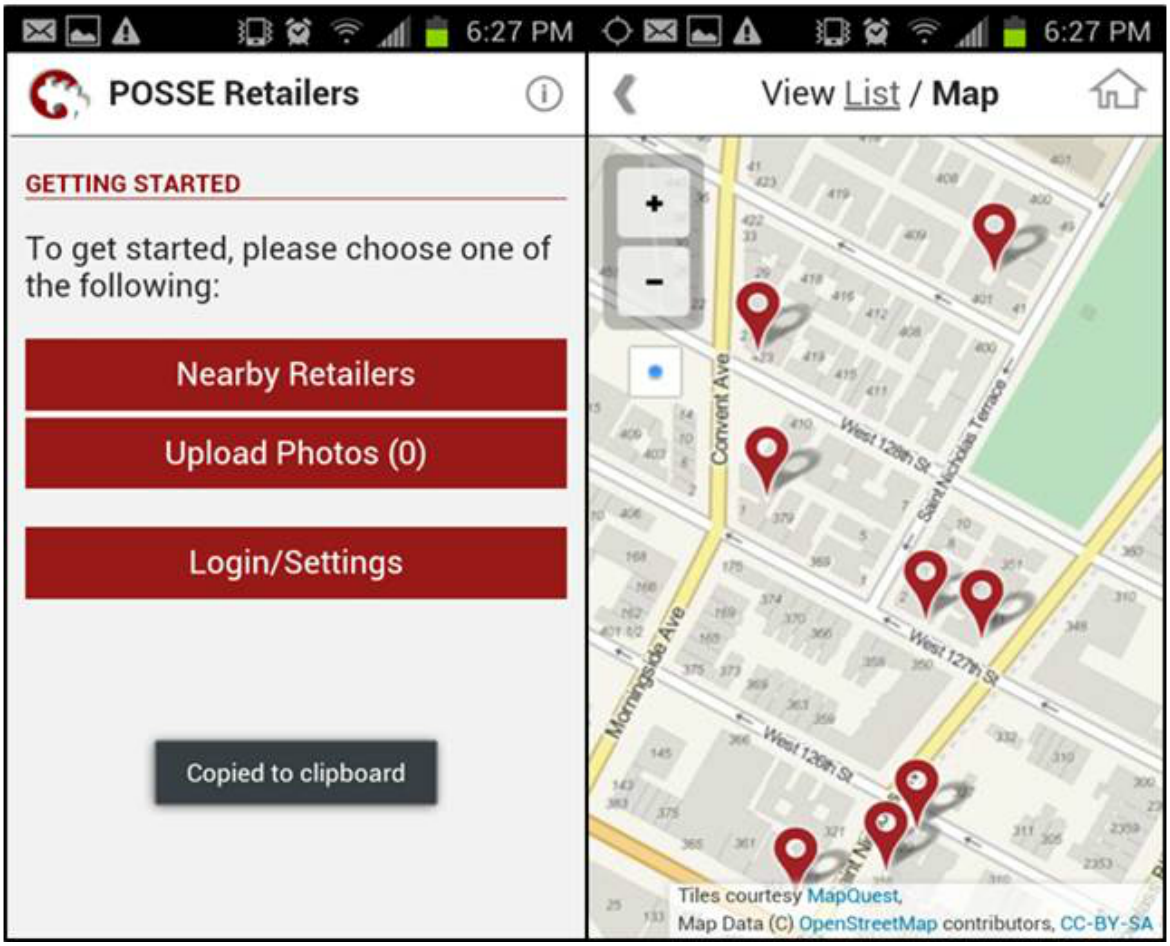
Table 1. Principles and examples for streamlining user interface components for the IVR interface.

Principle	Issues	Solutions
Simplicity	Audio surveys via IVR may take much longer than pen-and-paper surveys and can make complicated survey questions difficult.	Utilize clear and simple language for survey questions; allow for shortcuts and skip patterns where feasible.
		Ideally, keep survey instruments short (approximately 5-10 minutes) and focused on a limited number of surveillance topics.
		Construct question items to be concise and direct. For surveys in both cities, items are allowed for primarily dichotomous (Boolean) or multiple-choice response categories for ease of data entry, although some items allowed surveyors to enter data (such as counts and prices).
		Test questions via IVR to clarify and simplify each item's question stem, response categories, and flow as experienced in the field.
Accuracy	Fieldworkers are speaking responses and cannot visibly see what is being recorded to check their entry, correct something they misspoke, or that the system misheard.	Incorporate a brief check after each question to assess whether the entry was entered correctly. Upon each answer, the system confirmed the response. For example, if a fieldworker entered "liquor" to the question "what type of store is this?," the IVR system would immediately say, "You said 'liquor.' Is that correct?" If the fieldworker entered "yes," they would move on to the next question but if they entered "no," the IVR system would ask the question again and allow the fieldworker to enter the correct answer.
	Background noise makes it hard for fieldworkers to hear survey questions and can also trigger false responses on the IVR if the system mistakes background noise for a response.	Design system to repeat questions until a response is provided.
		Use headphones and the phone's keypad to answer questions; mute phone in very loud areas.
Flexibility	With largely closed-ended responses, it was sometimes difficult to capture information on new brands appearing at the point-of-sale or other relevant commentary.	Allow surveyors to provide nonscripted information through IVR open-ended voice responses to specific questions, which may later be coded.
		Incorporate photos into the data collection process. Photos can provide qualitative data on product information that may not have been captured in the survey and can occasionally serve as a check on survey data.
	In the field, unexpected issues might arise, such as confrontation by a store employee, which require individuals to exit the store or hang up on the survey.	The survey should allow for fieldworkers to hang up midsurvey and pick up from the last section they had started without losing previous data entered for the store.
	Speaking responses into the phone to answer survey questions via IVR can call attention to fieldworkers, especially in small stores.	Fieldworkers should use headphones to listen to the survey and answer the questions using the phone's keypad, which works for most questions, with the exception of open-ended questions.

User Interface: Photographs

While the interface for taking and emailing photographs on a mobile phone is straightforward, these activities must be seamlessly integrated into a multicomponent, multimodal data collection workflow so as not to slow down the store assessment process. For the DC project, we established through field testing that only 1 exterior photograph of the main storefront and 1 interior photograph of a checkout countertop area were required. We encouraged fieldworkers to take additional photos to document tobacco advertising; in practice, however, fieldworkers could only email 1 photo at a time given the back-end need for linking photos to store IDs, which made the process time consuming and tedious. For the NYC project, we incorporated this learning from DC by building a basic mobile app that allowed fieldworkers to take and upload multiple photos for each store in a single step. Taking photographs and linking them to individual stores was streamlined within the app by allowing fieldworkers using GPS to view nearby stores in the "list" mode, which listed stores by proximity, or in the "map" mode, which displayed a map populated by nearby stores based on proximity (Figure 2). Fieldworkers could then select the store and take multiple photos, which were easily and automatically linked to each store.

Figure 2. Mobile app for NYC (map view).



Fieldworker Training and Technical Support

Once fieldworkers were trained on the basics of tobacco advertising, they tested the system with the project manager and in pairs. We established a Web forum during testing and training so that fieldworkers could describe issues that arose in field tests and the team could view daily posts and decide on solutions to common problems. Once protocols were established, fieldworkers underwent a comprehensive technical training with the project manager (Textbox 1).

Textbox 1. Technology training protocol.

- Conduct one-on-one instruction between project manager/trainer and fieldworker.
- Review how to use the IVR survey on project phones, how to access various project phone numbers, and how to turn on GPS tracking.
- In DC, review how to take appropriate photos with the project phone, attach the photo (or multiple photos) to an email, and send them to the designated project email addresses linked to the study database; in NYC, review how to use the application to identify stores and how to take and send photos.
- Practice taking the IVR survey in the office using photographs from the field to get comfortable with the survey flow.
- Project manager/trainer accompany each fieldworker independently into local neighborhood stores to practice using the technology in the "real world" setting of a retail outlet.
- After in-field training, project manager and fieldworker debrief to discuss challenges and strategies for how to deal with the technology in the retail environment.
- Provide booster trainings throughout project implementation for fieldworkers having difficulty with the technology.

We also established real-time technical support throughout implementation by ensuring a manager and/or technologist were on call when fieldworkers were working to resolve problems quickly. For example, during the DC project testing, the IVR provider would occasionally shut down. On these occasions, fieldworkers would contact the person on call, who could quickly resolve the issue by notifying the IVR provider and keeping fieldworkers updated.

Back-End Data Systems

To link store, survey, and photo data in the DC project, the project manager established store IDs linked to store names and addresses. Fieldworkers then entered the IDs when conducting

store surveys or emailing photographs. For NYC, the consistent use of accurate store IDs proved challenging among the volunteer fieldworker staff. Because the photo data were linked to store names and addresses through the mobile app, we established a system of linking survey data from the phone surveys and photo data from the mobile app via a timestamp. Figure 3 provides a visual of the ID linking process for each project. Ideally, the linking of survey and photo data to store IDs would be automatic via the use of a mobile app that incorporated survey, photo, and location data collection linked to a more advanced, integrated data infrastructure than was available at the time of these studies.

The full utilization of GPS technologies was limited in each of these projects given the early stages of systems development. For both projects, we conducted geocoding using a geographic information system (GIS) based on store address via a somewhat onerous process that required downloading and cleaning files, uploading files into GIS systems for geocoding, and downloading and cleaning again. For both cities, we used a minimum matching score of 90% to batch geocode the addresses and manually geocoded any matches below 90%. In a later version of the back-end system that was developed subsequent to the DC and NYC projects, all data systems were streamlined so that different sources of data—particularly GPS data based on fieldworker routing—store IDs, and data collected were synchronized automatically.

To develop the website, we worked closely with the technologist, testing different aspects of the site to specify the components of a simple and interactive dashboard that allowed staff to monitor data collection in real-time and to access updated datasets. Components included: a map of stores assessed by ZIP code (Figure 4); a table of individual store observations from the IVR survey updated in real time (Figure 5); photographs with specified identifiers (Figure 6); visuals of aggregate data for each variable in the survey updated in real-time (Figure 7); a copy of the survey and a test version of the survey; a project management calendar; and the ability to export survey and photographic data in standard file formats, such as comma separated values and Microsoft Excel.

Figure 3. Linkage process for store survey and photo data by project.

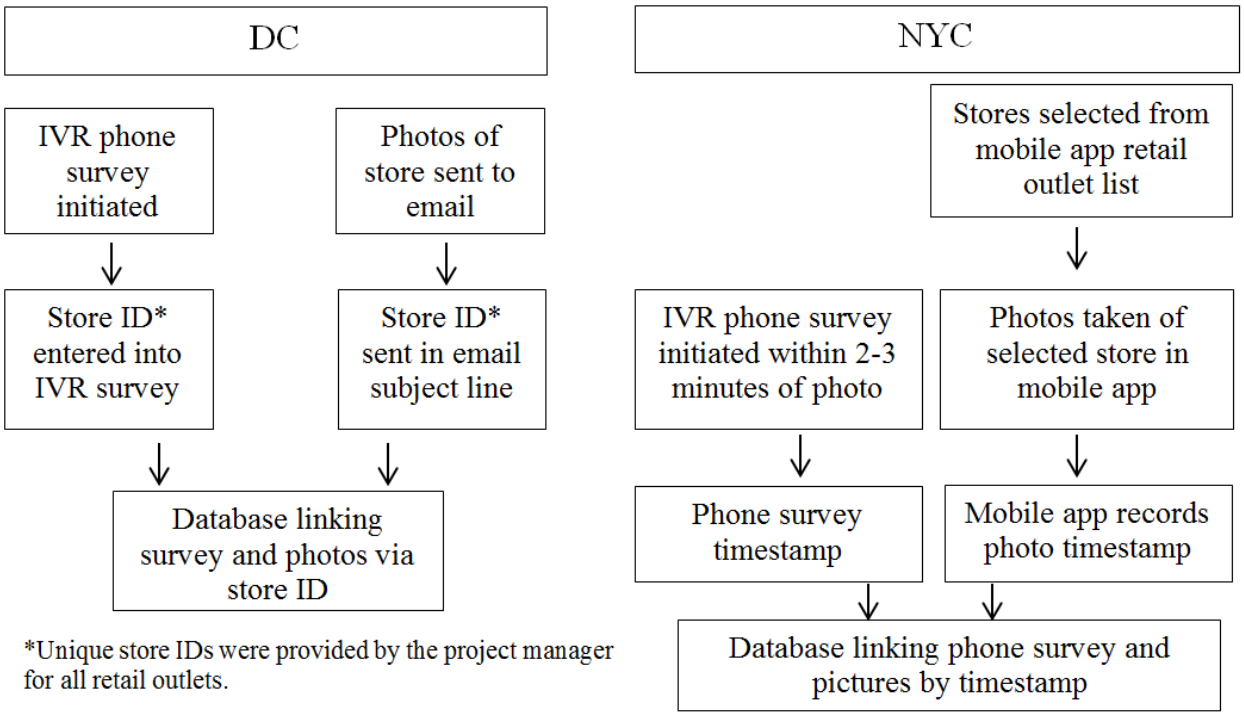


Figure 4. Back-end database with map of stores assessed by ZIP code.

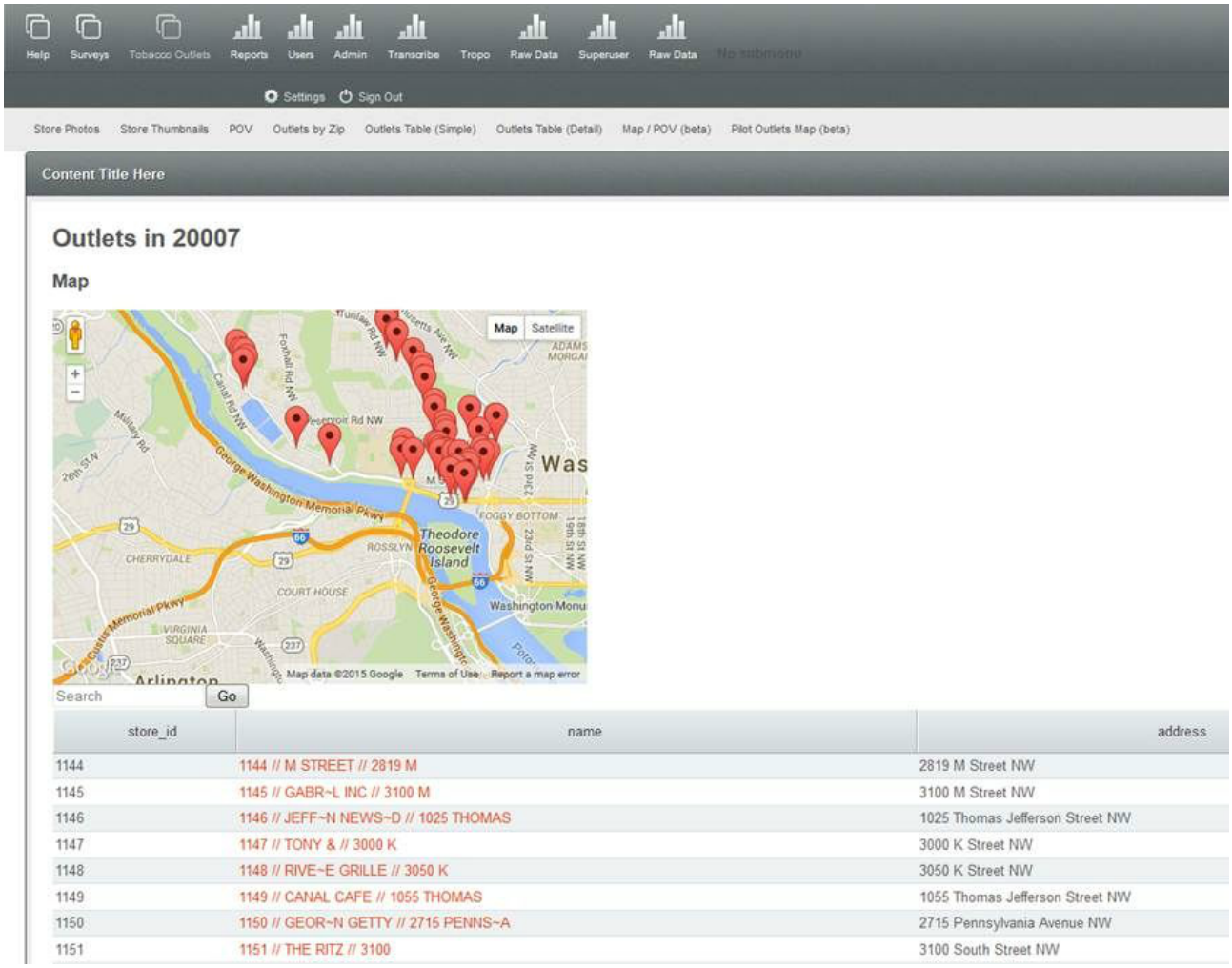


Figure 5. Back-End Database with Table of Individual Store Observations Updated in Real Time.

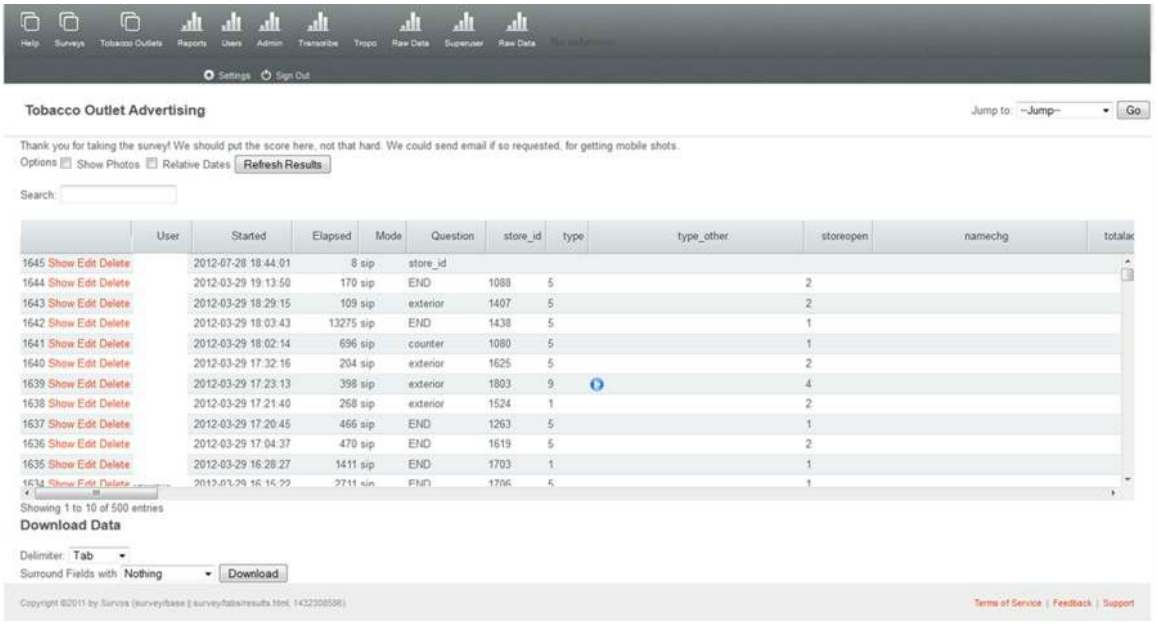


Figure 6. Back-End Database with Photographs from Data Collection.

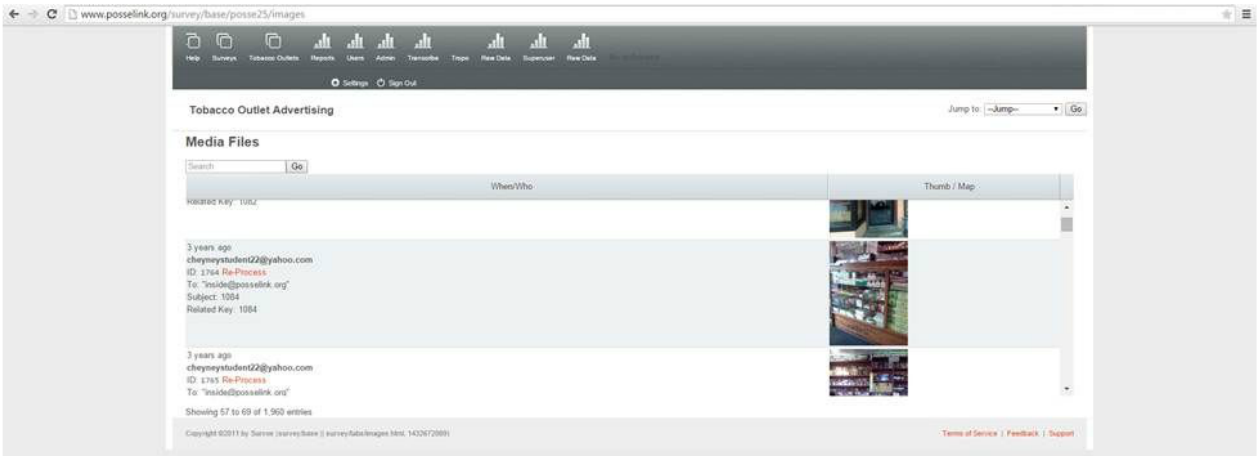
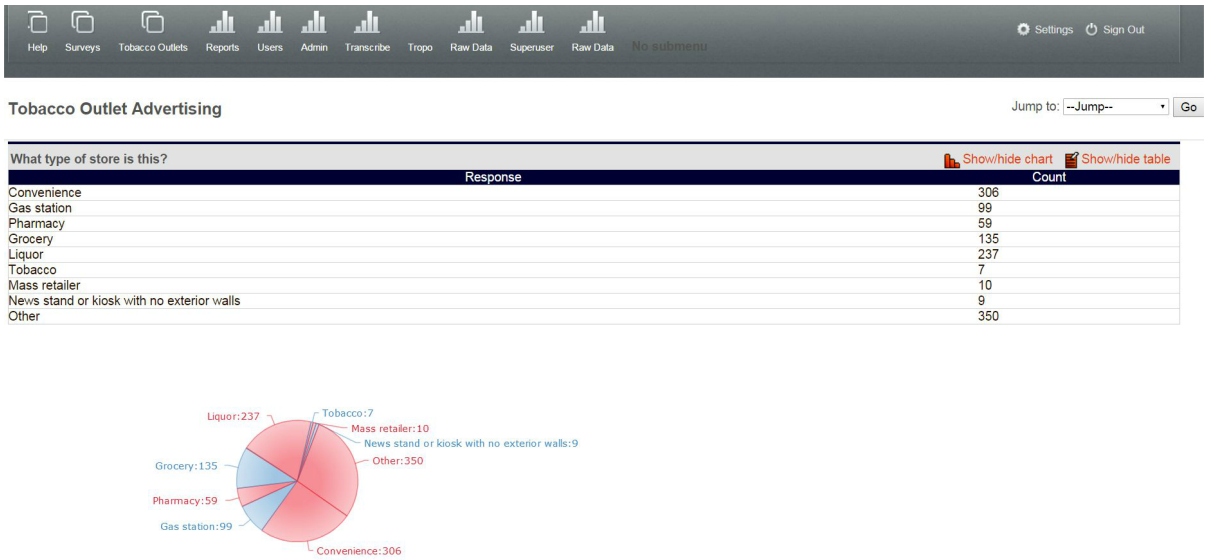


Figure 7. Back-End Database Summary Table and Pie Chart of Aggregated Variables.



Discussion

Multimodal mobile systems for point-of-sale surveillance present opportunities to collect, analyze, and disseminate retail data in real-time to researchers and community stakeholders. Yet implementation of seemingly straightforward tools requires that researchers and practitioners familiarize themselves with processes that are common to technology development and engineering [24-26]. Below we describe lessons learned from the DC and NYC projects that are applicable to mobile technology implementation projects for point-of-sale surveillance.

Lesson 1: Implementation of Mobile Systems Requires an Iterative Testing Process

Implementation of mobile systems introduces new technical decisions that can be resolved via an iterative testing approach prior to project launch. We utilized quick, continuous, and repeated testing to ensure that the technology, data collection workflow, and user-interface were workable and user-friendly in different situations and for multiple users. This process was critical to identifying technical, individual, and environmental challenges that influenced the performance of fieldworkers and the accuracy of data collection. Studies examining the use of mobile phones for interventional purposes have also noted the need to incorporate rapid, repeated testing and development phases in pilot work [26,27].

Testing time and a budget should be built into a pilot field testing period, with more investment required early on and generally decreasing costs as teams learn from each project. Budgets for pilot field testing mobile data collection will vary by project and may include fieldworker costs (unless volunteers are used), mobile phone costs (ie, phones, cellular and/or wireless service), technologist costs if an outside technology provider is used, and costs for the project management or research team. For these studies, Truth Initiative funded a local technology provider for the initial development of the multimodal mobile system and for support during the field testing and implementation process. Open source software or other resources to develop and implement mobile data collection surveillance studies are also becoming increasingly available [28].

Lesson 2: Workflow and User Interface Decisions Require Attention and Testing

Data collection workflow and the user interface take on increased importance with the use of new technologies for surveillance. Ease of use and usefulness to the end user is closely linked to the performance of a technology [29,30] and to effective implementation of mobile surveillance projects [31-34], thus it is critical to consider the experience of the primary users of the system. Therefore, as the main user of the system, we put the fieldworker at the center of the process for establishing the workflow protocol and streamlining the interface. This helped frame questions about what data to collect from a user-centered perspective of how the information could best be collected using the technology to ensure simplicity, accuracy, and flexibility.

Lesson 3: Training and Implementation Require Close Communication, Feedback, and Ongoing Technical Assistance

Developing feedback loops for consistent communication between fieldworkers, researchers, and the technologist—along with rapid ongoing technical assistance—is critical during testing and deployment to quickly resolve problems and keep field staff committed to using the technology correctly [34-36]. This also assures researchers that data collection and quality will not be undermined by technical glitches. During the iterative testing process, the Web forum and daily debriefs with fieldworkers kept all staff up to date with technical, workflow, or other field problems and allowed for immediate problem solving. This process served as a training ground for fieldworkers who were involved in the early stages of the project and, importantly, allowed for the development of knowledge and skills among all staff for quickly resolving technical glitches and moving the project forward.

Lesson 4: Back-End Data Infrastructure and Automated Systems Require Development and Integration

Automatic collection of data from the field to a back-end server provides many potential benefits, including allowing managers to utilize real-time quality control while data is being collected to improve data accuracy and quickly resolve logistical difficulties. However, these tools require that systems are integrated and, ideally, automatically link different sources of data from the field together for real-time monitoring, integration with GIS systems, and analyses. This process takes time and planning to identify the key issues for each project for integrating data sources, linking data from multimodal mobile sources to store identifiers, ensuring GPS can be collected and used efficiently in GIS systems, and ensuring skilled staff or outside consultants are available to develop and maintain this automated back-end infrastructure. Careful planning for back-end systems can ensure that a key benefit of multimodal mobile surveillance—rapid, automated, and visualized data collection—can be fully utilized.

Limitations

Our findings should be considered with the following limitations:

1. This study was not designed to examine the validity or reliability of mobile phone data collection for surveillance compared with more traditional paper-based methods, so it cannot be assumed these methods lead to improved data quality compared with paper surveys. We have, however, clarified some of the benefits of mobile data collection in comparison with paper surveys.
2. Fieldworkers and staff used their own phones in the early stages of testing until phone criteria were developed and phones were purchased. In projects where staff cannot use personal phones, field testing to establish mobile phone criteria would likely require additional funding or rely more on online research. Given that mobile technology is advancing at a rapid pace, projects that last more than 6 months to 1 year may need to update mobile phone hardware to take advantage of new mobile phone technologies as well as developments in software and automatic data integration, all of which will impact multiple components of a project.
3. Lists of licensed tobacco retail outlets in the geographic areas surveyed were available for these projects. Not all cities or states require or provide such lists, thus field surveying of stores may be necessary in some areas and may require additional strategies for identifying retail outlets and linking data from different sources to specific stores through mobile systems.

These projects were conducted in the United States, thus different or additional lessons may be applicable in international settings. Further, as technology continues to evolve, new lessons may apply.

Conclusions

Mobile technologies may improve surveillance efforts and facilitate rapid policy responses to emerging tobacco products and industry marketing practices. Information on technology implementation in health care and health research is sparse, which may be one factor in the limited number of large-scale implementation efforts despite a growing literature indicating promising findings from technology and health pilot studies [37]. Implementing new technologies with an iterative testing framework that (1) ensures a smooth workflow and user-friendly interfaces; (2) allows for training, ongoing communication, and feedback among all staff; and (3) supports infrastructure for back-end data integration is critical for successful system performance.

This is the first study to examine implementation needs for conducting multimodal mobile tobacco retail surveillance. Although mobile technologies are evolving rapidly, these implementation lessons are relevant for those who want to harness the strengths of any new mobile system for point-of-sale surveillance or collection of other types of neighborhood observational data. Findings suggest the need for further research examining mobile surveillance implementation challenges, as well as funding for developing and implementing systems that allow public health practitioners and researchers to improve surveillance by taking advantage of the communications tools of the twenty-first century.

Acknowledgments

This research was supported by the Centers for Disease Control and Prevention, Communities Putting Prevention to Work from the District of Columbia Department of Health (contract PO358719 to TRK), and Truth Initiative (known as the American Legacy Foundation at the time of this study) We thank Denise Grant and Bonita McGee from the DC Department of Health.

Conflicts of Interest

One of the authors, Michael Tacelosky, is the owner of the local information technology service provider, Survos Observe, used to develop and support the multimodal surveillance system.

References

1. World Health Organization. Immunization Surveillance, Assessment and Monitoring. Geneva, Switzerland: World Health Organization; 2014. URL: http://www.who.int/immunization/monitoring_surveillance/en/ [accessed 2014-12-29] [WebCite Cache ID 6VBu8AiEb]
2. Cohen JE, Planinac LC, O'Connor SC, Lavack AM, Robinson DJ, Thompson FE. Keeping the point-of-sale environment at the forefront. Am J Public Health 2008 Jan;98(1):5-6. [doi: 10.2105/AJPH.2007.124321] [Medline: 18048773]
3. Tomlinson M, Solomon W, Singh Y, Doherty T, Chopra M, Ijumba P, et al. The use of mobile phones as a data collection tool: A report from a household survey in South Africa. BMC Med Inform Decis Mak 2009;9(1):51 [FREE Full text] [doi: 10.1186/1472-6947-9-51] [Medline: 20030813]
4. Kanter R, Alvey J, Fuentes D. A novel mobile phone application to assess nutrition environment measures in low- and middle-income countries. Food Nutr Bull 2014 Sep;35(3):296-300. [Medline: 25902589]
5. Cohen DA, Schoeff D, Farley TA, Bluthenthal R, Scribner R, Overton A. Reliability of a store observation tool in measuring availability of alcohol and selected foods. J Urban Health 2007 Nov;84(6):807-813 [FREE Full text] [doi: 10.1007/s11524-007-9219-x] [Medline: 17763963]
6. Lee Joseph GL, Henriksen L, Myers AE, Dauphinee AL, Ribisl KM. A systematic review of store audit methods for assessing tobacco marketing and products at the point of sale. Tob Control 2014 Mar;23(2):98-106 [FREE Full text] [doi: 10.1136/tobaccocontrol-2012-050807] [Medline: 23322313]
7. Morland K, Wing S, Diez RA, Poole C. Neighborhood characteristics associated with the location of food stores and food service places. Am J Prev Med 2002 Jan;22(1):23-29. [Medline: 11777675]
8. Campaign for Tobacco-Free Kids. Store Alert: Retail Store Survey URL: http://www.storealert.org/default_flash.asp [accessed 2014-12-29] [WebCite Cache ID 6VBqzSJ61]
9. Glanz K, Sutton NM, Jacob AKR. Operation storefront Hawaii: Tobacco advertising and promotion in Hawaii stores. J Health Commun 2006;11(7):699-707. [doi: 10.1080/10810730600941471] [Medline: 17074736]
10. Gong T, Lv J, Liu Q, Ren Y, Li L, Kawachi I, Community Interventions for Health (CIH) Collaboration. Audit of tobacco retail outlets in Hangzhou, China. Tob Control 2013 Jul;22(4):245-249. [doi: 10.1136/tobaccocontrol-2011-050038] [Medline: 22180327]
11. Rose SW, Myers AE, D'Angelo H, Ribisl KM. Retailer adherence to Family Smoking Prevention and Tobacco Control Act, North Carolina, 2011. Prev Chronic Dis 2013;10:E47 [FREE Full text] [doi: 10.5888/pcd10.120184] [Medline: 23557638]
12. Salloum RG, Nakkash RT, Myers AE, Wood KA, Ribisl KM. Point-of-sale tobacco advertising in Beirut, Lebanon following a national advertising ban. BMC Public Health 2013;13:534 [FREE Full text] [doi: 10.1186/1471-2458-13-534] [Medline: 23731766]
13. Spanopoulos D, Ratschen E, McNeill A, Britton J. Retail price and point of sale display of tobacco in the UK: A descriptive study of small retailers. PLoS One 2012;7(1):e29871 [FREE Full text] [doi: 10.1371/journal.pone.0029871] [Medline: 22242183]
14. Cantrell J, Kreslake JM, Ganz O, Pearson JL, Vallone D, Anesetti-Rothermel A, et al. Marketing little cigars and cigarillos: Advertising, price, and associations with neighborhood demographics. Am J Public Health 2013 Oct;103(10):1902-1909. [doi: 10.2105/AJPH.2013.301362] [Medline: 23948008]
15. Kirchner T, Cantrell J, Anesetti-Rothermel A, Pearson J, Cha S, Kreslake J, et al. Individual mobility patterns and real-time geo-spatial exposure to point-of-sale tobacco marketing. New York: ACM; 2012 Presented at: ACM Wireless Health'12; Oct 23-25, 2012; San Diego, CA. [doi: 10.1145/2448096.2448104]
16. Kirchner TR, Cantrell J, Anesetti-Rothermel A, Ganz O, Vallone DM, Abrams DB. Geospatial exposure to point-of-sale tobacco: Real-time craving and smoking-cessation outcomes. Am J Prev Med 2013 Oct;45(4):379-385 [FREE Full text] [doi: 10.1016/j.amepre.2013.05.016] [Medline: 24050412]
17. Cantrell J, Ganz O, Anesetti-Rothermel A, Harrell P, Kreslake JM, Xiao H, et al. Cigarette price variation around high schools: Evidence from Washington DC. Health Place 2015 Jan;31:193-198. [doi: 10.1016/j.healthplace.2014.12.002] [Medline: 25560754]

18. U.S. Census Bureau. American Community Survey URL: http://www.census.gov/acs/www/ [accessed 2014-12-29] [WebCite Cache ID 6VBtnuJcI]
19. New York City Department of City Planning. Manhattan Community District 10. New York: New York City Department of City Planning; 2010. URL: http://www.nyc.gov/html/dcp/pdf/lucds/mn10profile.pdf [accessed 2014-12-29] [WebCite Cache ID 6YoncKJ11]
20. New York City Department of City Planning. Census Demographic Maps. New York: New York City Department of City Planning; 2013. URL: http://www.nyc.gov/html/dcp/html/census/demo_maps_2010.shtml [accessed 2015-05-26] [WebCite Cache ID 6YonfK6sW]
21. Ganz O, Cantrell J, Moon-Howard J, Aidala A, Kirchner TR, Vallone D. Electronic cigarette advertising at the point-of-sale: A gap in tobacco control research. Tob Control 2015 Mar;24(e1):e110-e112. [doi: 10.1136/tobaccocontrol-2013-051337] [Medline: 24619455]
22. Office of the Chief Technology Officer, District of Columbia. District of Columbia: Data Catalog. 2012. URL: http://opendata.dc.gov/ [accessed 2015-07-31] [WebCite Cache ID 6aQAOo2e5]
23. Environmental Systems Research Institute. ArcGIS Desktop: Advanced 10.1 ed. Redlands, CA: Envrionmental Systems Research Institute; 2012.
24. Boehm BW. A spiral model of software development and enhancement. Computer 1988 May;21(5):61-72. [doi: 10.1109/2.59]
25. Hartson H, Hix D. Toward empirically derived methodologies and tools for human-computer interface development. Int J Man-Mach Stud 1989 Oct;31(4):477-494. [doi: 10.1016/0020-7373(89)90005-9]
26. Ben-Zeev D, Schueller SM, Begale M, Duffecy J, Kane JM, Mohr DC. Strategies for mHealth research: Lessons from 3 mobile intervention studies. Adm Policy Ment Health 2015 Mar;42(2):157-167. [doi: 10.1007/s10488-014-0556-2] [Medline: 24824311]
27. Bergstrom RJ. Web Survey and Forms Usability Design and Testing. 2014. URL: http://www.slideshare.net/JenniferRomanoBergstrom/web-survey-and-forms-usability-design-testing [accessed 2015-05-26] [WebCite Cache ID 6YomWXrNA]
28. CommCare. URL: https://www.commcarehq.org/home/ [accessed 2015-05-26] [WebCite Cache ID 6YonUZYEr]
29. Karsh B. Beyond usability: Designing effective technology implementation systems to promote patient safety. Qual Saf Health Care 2004 Oct;13(5):388-394 [FREE Full text] [doi: 10.1136/qhc.13.5.388] [Medline: 15465944]
30. Constantine LL, Lockwood LA. Software for Use: A Practical Guide to the Models and Methods of Usage-Centered Design. Boston, MA: Pearson Education; 1999.
31. Berg M, Wariero J, Modi V. Every Child Counts: The Use of SMS in Kenya to Support the Community Based Management of Acute Malnutrition and Malaria in Children Under Five. 2009. URL: http://www.cmamforum.org/resource/305 [accessed 2015-05-26] [WebCite Cache ID 6Yom9OapF]
32. World Health Organization. mHealth: New Horizons for Health Through Mobile Technologies. Geneva, Switzerland: WHO Press; 2011. URL: http://apps.who.int/iris/bitstream/10665/44607/1/9789241564250_eng.pdf [accessed 2015-11-17] [WebCite Cache ID 6d6jPkwmp]
33. Rajatonirina S, Heraud JM, Randrianasolo L, Orelle A, Razanajatovo NH, Raoelina YN, et al. Short message service sentinel surveillance of influenza-like illness in Madagascar, 2008-2012. Bull World Health Organ 2012 May 1;90(5):385-389 [FREE Full text] [doi: 10.2471/BLT.11.097816] [Medline: 22589573]
34. Brinkel J, Krämer A, Krumkamp R, May J, Fobil J. Mobile phone-based mHealth approaches for public health surveillance in sub-Saharan Africa: A systematic review. Int J Environ Res Public Health 2014 Nov;11(11):11559-11582 [FREE Full text] [doi: 10.3390/ijerph111111559] [Medline: 25396767]
35. Pascoe L, Lungo J, Kaasbøll J, Koleleni I. Collecting integrated disease surveillance and response data through mobile phones. In: Proceedings of the IST-Africa 2012 Conference and Exhibition. 2012 Presented at: IST-Africa 2012 Conference and Exhibition; 2012; Dar es Salaam, Tanzania.
36. Kamanga A, Moono P, Stresman G, Mharakurwa S, Shiff C. Rural health centres, communities and malaria case detection in Zambia using mobile telephones: A means to detect potential reservoirs of infection in unstable transmission conditions. Malar J 2010;9:96 [FREE Full text] [doi: 10.1186/1475-2875-9-96] [Medline: 20398318]
37. Nhavoto JA, Grönlund A. Mobile technologies and geographic information systems to improve health care systems: A literature review. JMIR Mhealth Uhealth 2014;2(2):e21 [FREE Full text] [doi: 10.2196/mhealth.3216] [Medline: 25099368]

Abbreviations

DC: District of Columbia
GIS: geographic information system
GPS: global positioning system
IVR: interactive voice response
NYC: New York City

Edited by G Eysenbach; submitted 31.12.14; peer-reviewed by A Chan, S Regan; comments to author 21.04.15; revised version received 29.05.15; accepted 07.10.15; published 26.11.15.

Please cite as:
Cantrell J, Ganz O, Ilakkuvan V, Tacelosky M, Kreslake J, Moon-Howard J, Aidala A, Vallone D, Anesetti-Rothermel A, Kirchner TR
Implementation of a Multimodal Mobile System for Point-of-Sale Surveillance: Lessons Learned From Case Studies in Washington, DC, and New York City
JMIR Public Health Surveill 2015;1(2):e20
URL: http://publichealth.jmir.org/2015/2/e20/
doi:10.2196/publichealth.4191
PMID:27227138

Original Paper

Agenda Setting for Health Promotion: Exploring an Adapted Model for the Social Media Era

Yousef Albalawi[1,2*], MBBS, MSc Public Health; Jane Sixsmith[1*], RN, HV, BSc, MA, PhD

[1]Health Promotion Research Centre, National University of Ireland Galway, Galway, Ireland

[2]Public Health Administration, Ministry of Health, Medina, Saudi Arabia

[*]all authors contributed equally

Corresponding Author:
Yousef Albalawi, MBBS, MSc Public Health
Health Promotion Research Centre
National University of Ireland Galway
University Road
Galway,
Ireland
Phone: 353 83 1232203
Fax: 353 91 494577
Email: y.albalawi@gmail.com

Abstract

Background: The foundation of best practice in health promotion is a robust theoretical base that informs design, implementation, and evaluation of interventions that promote the public's health. This study provides a novel contribution to health promotion through the adaptation of the agenda-setting approach in response to the contribution of social media. This exploration and proposed adaptation is derived from a study that examined the effectiveness of Twitter in influencing agenda setting among users in relation to road traffic accidents in Saudi Arabia.

Objective: The proposed adaptations to the agenda-setting model to be explored reflect two levels of engagement: agenda setting within the social media sphere and the position of social media within classic agenda setting. This exploratory research aims to assess the veracity of the proposed adaptations on the basis of the hypotheses developed to test these two levels of engagement.

Methods: To validate the hypotheses, we collected and analyzed data from two primary sources: Twitter activities and Saudi national newspapers. Keyword mentions served as indicators of agenda promotion; for Twitter, interactions were used to measure the process of agenda setting within the platform. The Twitter final dataset comprised 59,046 tweets and 38,066 users who contributed by tweeting, replying, or retweeting. Variables were collected for each tweet and user. In addition, 518 keyword mentions were recorded from six popular Saudi national newspapers.

Results: The results showed significant ratification of the study hypotheses at both levels of engagement that framed the proposed adaptions. The results indicate that social media facilitates the contribution of individuals in influencing agendas (individual users accounted for 76.29%, 67.79%, and 96.16% of retweet impressions, total impressions, and amplification multipliers, respectively), a component missing from traditional constructions of agenda-setting models. The influence of organizations on agenda setting is also highlighted (in the data of user interactions, organizational accounts registered 17% and 14.74% as source and target of interactions, respectively). In addition, 13 striking similarities showed the relationship between newspapers and Twitter on the mentions trends line.

Conclusions: The effective use of social media platforms in health promotion intervention programs requires new strategies that consider the limitations of traditional communication channels. Conducting research is vital to establishing a strong basis for modifying, designing, and developing new health promotion strategies and approaches.

(JMIR Public Health Surveill 2015;1(2):e21) doi:10.2196/publichealth.5014

KEYWORDS

agenda setting, health promotion, social media, Twitter, health communication, Saudi Arabia, road traffic accidents

Introduction

Background

Communication is a core component of many effective health promotion interventions and change processes at individual and community levels [1]. In the social media age, the emergence of eHealth communication is expected to significantly enhance the efficacy of health promotion programs. The evolution of social media stimulated a shift of the communication equation from a top-down, expert-to-consumer approach to a nonhierarchical, dialog-based strategy. Consequently, communication has become an individual and community enabler in terms of achieving development goals, including health development [2]. Korda [3] indicated that an important characteristic of Web-based interventions is the sense of empowerment that it endows people and groups as they make decisions related to health; this feature is a positive influence on communities and individuals who are actively aiming for healthy behaviors and lifestyle changes. With these considerations in mind, we investigated the use of the agenda-setting function of health promotion interventions in the social media era. Specifically, we examined the effectiveness of Twitter as a social media platform in influencing agenda setting among users in relation to road traffic accidents in Saudi Arabia.

Road Traffic Accidents

Globally, road traffic accidents result in 1.24 million deaths and 20 to 50 million injuries per year, many of which cause permanent disabilities [4]. In Saudi Arabia, the 544,000 yearly accidents cause 7153 fatalities and more than 39,000 injuries [5]. Eighty-one percent of deaths in Ministry of Health hospitals are the result of road traffic accidents [6]. The World Health Organization recommendations emphasize the consideration of road safety as a public health issue [7], with a focus on persuading policy and decision makers to place road traffic accidents on their agendas as a major problem and implement measures for improving related interventions.

Maximizing the effectiveness of social media for the promotion and protection of health necessitates intervention programs based on a thorough scientific understanding of how communication and media action theories and models are prioritized [8,9]. Agenda-setting theory has been examined within the sphere of social media and shows promise for the promotion of effective health practices [10,11,12].

According to Kaplan and Haenlein, social media is “a group of Internet-based applications that build on the ideological and technological foundations of Web 2.0, and that allow the creation and exchange of User Generated Content” [13]. This definition covers many types of social media including Twitter, Facebook, and Instagram. These platforms have powerful characteristics which make them effective channels for communication-based activities. An interesting development in recent years is the significant increase in the availability of social media; this growth is expected to continue [14].

The development of social media has been recognized as an opportunity for the promotion of the public’s health demonstrated through the concept of infodemiology, a term coined by Eysenbach [15]. Infodemiology is the melding of health informatics and epidemiology and has been defined as “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” [16,17].

Infodemiology is based on the idea that the vast quantities of communication data generated by social media can be used for public health [16]. We live in a digital world where people communicate using Internet channels supported by highly advanced technologies. These communication channels are characterized by an ability to track activities and collect information and data about them. For example, social media platforms generate data that reflect people’s behaviors and record, in real time, large parts of their daily life, including their health status [16,18]. When suitable metrics and measures are applied, these data can provide valuable information that can inform policies, strategies, and decisions for public health at the level of policy makers and of the population [17].

These data provide a new level of information that was not measurable before this era [17]. Currently, only a small proportion will be analyzed (in 2013, only 5% of these data were analyzed [19]) due to a lack of methods and measures for collecting, analyzing, and interpreting such data [16,17,20]. Nevertheless, infodemiology advances our understanding and provides methods that can move public health to a new level of practice and research [16,17]. Applications of infodemiology can harmonize the research and practice of public health through the analysis of so-called “big data” in the era of social media [16,17]. Examples of infodemiology applications include tracking user activities on microblogging platforms such as Twitter [16]. This study explores user activities on Twitter in relation to public health and as such can be positioned in the context of infodemiology.

Twitter

Twitter is “an information network made up of 140-character messages called Tweets” [21]. It is a social and microblogging service that enables participants to post messages and follow others’ posts. Outside China, 53% of the Internet population has Twitter accounts, and 69% of online adults browse Twitter [22]. The 2015 statistics for Saudi Arabia show that in a population of 28 million, more than 18 million are Internet users [23], 60% of whom have Twitter accounts and 33% of Internet users are active Twitter users [22]. Apart from being among the top-ranked countries in terms of registered users, Saudi Arabia is number one globally in terms of visitation rates (logged-out users) [22].

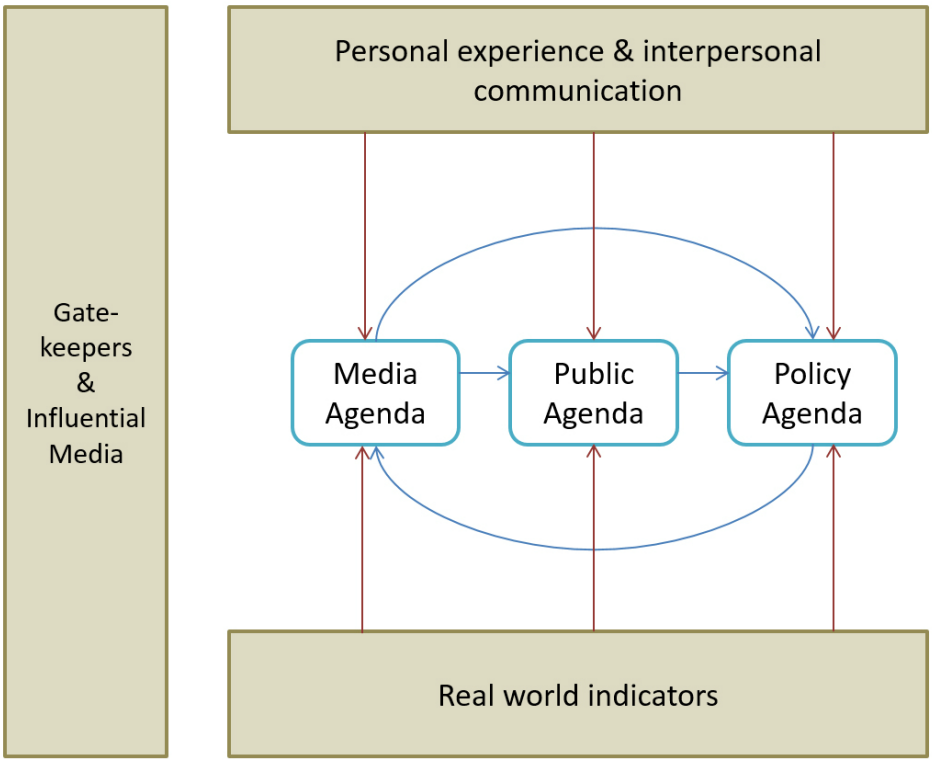
Agenda-Setting Theory

Lippman [24] first expressed the idea of agenda setting, which was subsequently developed by Lasswell [25] and Cohen [26], culminating in agenda-setting theory through the work of McCombs and Shaw [27]. The core concept of agenda setting assumes that media stimulates the awareness of people regarding certain issues. This assumption is grounded on two main principles: (1) media shapes and filters reality before presenting

it to people and (2) these channels determine the priority with which individuals regard salient issues [28]. Rogers and Dearing [29] proposed an agenda-setting model that comprises three components: media agenda, public agenda, and policy agenda. Each of these agendas represents issues that are the chief concerns of a particular stakeholder. The interrelationship among these components forms the core of agenda-setting theory [30]. Figure 1 shows the process of agenda setting among the three main components according to Rogers and Dearing's model [29].

As indicated in the model, media agenda setting refers to traditional media organization decisions on which issues to discuss through their channels. Public agenda setting revolves around the issues that are considered important to the general public. Policy agenda setting involves official organizations or government agencies that determine which issues are important and worthy of discussion [31].

Figure 1. The three main components of the agenda-setting model.



Agenda Setting for Health Promotion

Kozel et al [32] developed agenda setting in the context of public health and health promotion through the process of health promotion agenda-setting [12,32,33,34]. Agenda setting is about the interrelationship of the domains of policy, media, and public agendas; health promotion agenda-setting is about how health issues move through agendas to the point that they become actionable by policymakers [35]. Health promotion agenda-setting shifts the focus from the traditional health education target of individual risk behavior change to the formulation and adoption of innovative health policies which advocate for the public's health at population level [12,32]. Kozel et al [12,34], in response to an identified gap—the omission of agenda setting from health promotion planning models relating to innovation and diffusion—have developed a model of the health promotion agenda-setting process. The construction of this model includes the interrelated constructs of the media, policy, and public agendas with the integration of the seven responsibilities of health educators: assessment, planning, implementation, coordination, evaluation, acting as a resource person, and advocating for health [34]. Through the development of health promotion agenda-setting, including lessons learned from its practical application, a range of factors has emerged that enhances the diffusion of health promotion and disease prevention innovations [33]. These include characteristic factors such as demographic descriptors; design factors such as strategies and methods used; and mechanism factors such as shared vision, synchronicity, salience, and social justice [33]. Kozel et al [33] identify ten key activities for agenda setters to use in practice, two of which are tailoring strategies to prioritize a health issue in a population and sustaining salience of an issue in the domains of policy, media, and public agendas. The application of health promotion agenda-setting in practice enables a comprehensive, planned, innovative, and sustainable course of action which facilitates prioritization of public health problems and the identification of alternative solutions [12]. Health promotion agenda-setting contributes to health promotion leadership and provides a mechanism through which to improve the formulation and adoption of health policy.

In addition to the work by Kozel and colleagues on the development and application of health promotion agenda-setting, the concept and components of agenda setting have been used in public health and health promotion in a range of areas [36-42]. Understanding, researching, and implementing the use of agenda

setting for health promotion practice will improve its performance and boost intervention outcomes [12]. This is particularly important in the era of social media, a relatively new addition to the media landscape that warrants further exploration in the context of agenda setting.

Agenda Setting in the Social Media Era

Agenda-setting initiatives have been extensively studied and developed by researchers and practitioners. New frames and models have been proposed with emphasis placed on the ideal match between changes people and societies are undergoing in the social media era and agenda setting for public health [43]. Given that social media was nonexistent during the introduction and early development of agenda-setting theory, this has not been comprehensively investigated in previous research [44].

Simple application of agenda setting in the era of social media does not reflect the complex process of communication resulting from the use of social media platforms [44]. We argue that understanding agenda-setting theory in the social media era should cover two levels of engagement: the first centers on agenda setting within the social media sphere and the second is related to the position of social media within the classic agenda-setting process implemented in the real world. Here we propose social media in the agenda-setting context as an independent body governed by its own agenda..

Social Media Agenda Setting

As proposed by Dearing and Rogers [45], agenda setting is best understood as a process of interaction; it therefore revolves around the flow of agendas from one component to another. Within the agenda-setting process, an important task is to identify who owns specific agendas and who interacts with other stakeholders.

Individual Agenda

Social media offers numerous platforms where people can communicate and interact. One of the most important changes in agenda setting within the social media realm is the shift in power towards the public in terms of control over communication; this shift was triggered by the fact that with social media technologies, individuals become active producers instead of functioning merely as receivers of information. Bekkers et al [46] argue that Web 2.0 has shifted political mobilization from a traditional mass-oriented movement to one driven by individuals and small groups of people.

Furthermore, individuals differ significantly in how they respond to the media agenda [47]. The power that individuals have gained in the social media era enables them to directly communicate their arguments, opinions, and agendas to the world. Supported by highly interactive features and user-generated content, social media platforms allow individuals to control what they receive, from whom, and how much according to their interests [48]. Tran and Johnson [49] claim that one of the opportunities provided by social media to agenda-setting research is the empowerment of individuals in developing their personal agendas [49]. We argue that such opportunity extends not only to development but also to influence over agendas. In real communities, an individual is a member of the public and thus can adopt and influence a given agenda advocated by a specific community. Similarly, an individual can hold membership in any organization and assist this organization by adopting and influencing its agenda. In classic agenda setting, individuals are always regarded as members of the public because influence is acquired through a process called “agenda melding” [39]. Although the emergence of social media does not cancel the role of agenda melding, it may extend the role of individuals by assigning them effective positions within the social process. This perspective is supported by the findings of Althaus and Tewksbury [50] and Conway and Patterson [51], who illustrated the differences in the power of individuals to control communication between traditional and social media.

Organizational Agenda

Social media has also redistributed the power to control communication at the organizational level. The nature of social media has allowed many types of organizations—not only media and policy institutions—to contribute to agenda setting. Similar to the shift in power at the public and individual level, changes at the organizational level have translated to organizational influence over and interaction with various agendas.

The effect of the presence of health organizations on these platforms has been explored in recent studies [52,53]. These studies include the examination of factors such as those associated with the organizations’ ability to engage and measures that directly affect the organizations’ influence [53].

The organizational agenda is not a new concept. Berger [31] pointed out that organizations are effective agenda-setting actors that can establish agenda through funding, lobbying, and advertising, thereby influencing the specific issues that are discussed in societies.

The authors propose to regard organizations as essential stakeholders in agenda setting because they can interact with different community actors, including the public, media, and policy makers.

A New Contributor to the Agenda-Setting Process

As previously stated, the nature of social media with its two-way communication platforms and channels differs completely from that of traditional one-way mass communication channels. The social media age has driven changes in the manner by which information is disseminated. This era has decentralized traditional communication, thereby diminishing its power in shaping the issues that people think about [49]. Researchers have examined the relationship between traditional media (eg, newspapers and television) and social media (eg, Twitter and YouTube, a video-sharing website) [48,54]. The findings suggest that the social media realm is an independent arena that can affect and be affected by traditional media [54]. Research confirms traditional media’s influence over the social media agenda and vice versa [45,46,55].

McCombs [47] and Meraz [56], among others, have highlighted the manner by which social media influences agenda setting within the traditional media realm. An interesting finding is that the influence of social media not only covers the traditional

media agenda but also extends to public and policy agendas [49]; these new channels affect the entire agenda-setting process. In exploring the relationship between social media and other agenda-setting components, many researchers distinguish social media from traditional media. An example is the separate examination of social media's influence on public and policy agendas [57]. McCombs [47], who pointed out that social media redesigned agenda setting, supports this approach by adding a new contributor (ie, social media) to the process.

Collectively, the aforementioned studies focused not only on the discrepancy between traditional and social media in their effectiveness as communication tools but also on the independence of social media as an agenda-setting channel. Their findings suggest that studying social media as part of traditional media in the agenda-setting context is an unsuitable framework from which to understand the complexity of the agenda-setting process within the arena occupied by modern media innovations. About 70% of journal articles that explored agenda setting in the social media age are concentrated in intermedia agenda setting between new and traditional media [49]. Yet, the findings on social media as an independent channel [55] lend support to the claim that social media redistributed the power of agenda setting by adding a new domain to classic agenda-setting theory. We argue that social media can be regarded as a separate body within the agenda-setting process, as ideas from this perspective have been previously put forward in the literature. Meraz, for example, proposed social media as a new component of agenda setting, although he treated the new channels within as traditional media [56]. On these grounds, we propose a model for agenda setting in the social media era that reflects two levels of engagement: agenda setting within the social media sphere and the position of social media within classic agenda setting. The capability of the proposed model was assessed on the basis of the hypotheses formulated in this work.

Study Hypotheses

Hypothesis 1 (H1) revolves around the agenda-setting process within social media and suggests a new model fitted to the uniqueness of agenda setting under a social media interface. H1 maintains that individual and organizational agendas constitute a new body of plans and schemes instead of falling within the category of media and policy agendas (see Figure 2).

To validate H1, we put forward the following subhypotheses:

> *H1-1a: Individual accounts are the most dominant accounts.*
>
> *H1-1b: Organizational accounts are more dominant than media and policy accounts.*

We used three measures to determine the validity of H1-1a and H1-1b: rank scale measures for retweet impressions, total impressions, and total amplification multipliers (see Table 1 for definitions of terms). Data on contributor interactions were used as support measures. Two measures were adopted to differentiate between public and individual personal agendas: the average of total impressions and the amplification multiplier. The total impressions indicates the accumulated number of times a tweet was received, and the average is a measure of how a single account can be an influential factor in agenda setting compared with other account types. We therefore propose an additional subhypothesis:

> *H1-1c: Individual accounts represent individual personal agendas in addition to public agendas.*

Table 1. Dataset variables.

Terminology	Definition
Retweets	Number of times a tweet is reposted or forwarded
Deliveries	Number of accounts to which a tweet is posted initially (equal to the number of followers the user has at that time)
Total impressions	Number of accounts that received the tweet; this includes direct post, retweets, and replies.
Retweet impressions	Number of impressions retweets of this tweet have generated
Amplification multiplier	The rate of amplification based on the tweet spread by retweets [(total exposure − impressions) / impressions] + 1

The data on contributor interactions were also used to determine the degree of influence of individual accounts as targets. We assume that when an account is targeted by other account types, these accounts represent individual agendas rather than public agendas. To examine nonaccount agendas, a critical requirement is determining that influence goes beyond accounts with special characteristics. For example, degree of influence is not restricted by a specific level of popularity. We thus propose H1-2:

> *H1-2: No correlation exists between account popularity and an account's degree of influence; that is, tweets that are extensively disseminated can be created by accounts with only a few followers.*

We used the correlation between two pairs of measures to test H1-2: the correlation between deliveries (number of times a tweet was received) and retweet impressions and that between impressions and the amplification multiplier.

Hypothesis 2 (H2) is related to the position of the social media agenda within classic agenda-setting theory. The theory posits that social media are new components incorporated into the three main elements of the classic agenda setting proposed by Rogers and Dearing [29] (Figure 3).

Determining the validity of H2 necessitates an investigation into the relationship between the social media agenda and the three other agenda types (media, public, and policy). However, the data collected in this study are useful only in exploring the

relationship between Twitter as a social media platform and newspapers as traditional media channels. The collected data also lack many of the characteristics required from evidentiary sources (ie, a 90-day data collection period is a short time frame.). Despite these limitations, the data can provide valuable insights into the interaction between Twitter and newspaper agendas.

H2 is articulated thus:

> *H2: The trend of social media mentions is similar to that of newspaper mentions.*

Figure 2. Adapted model of agenda setting within social media.

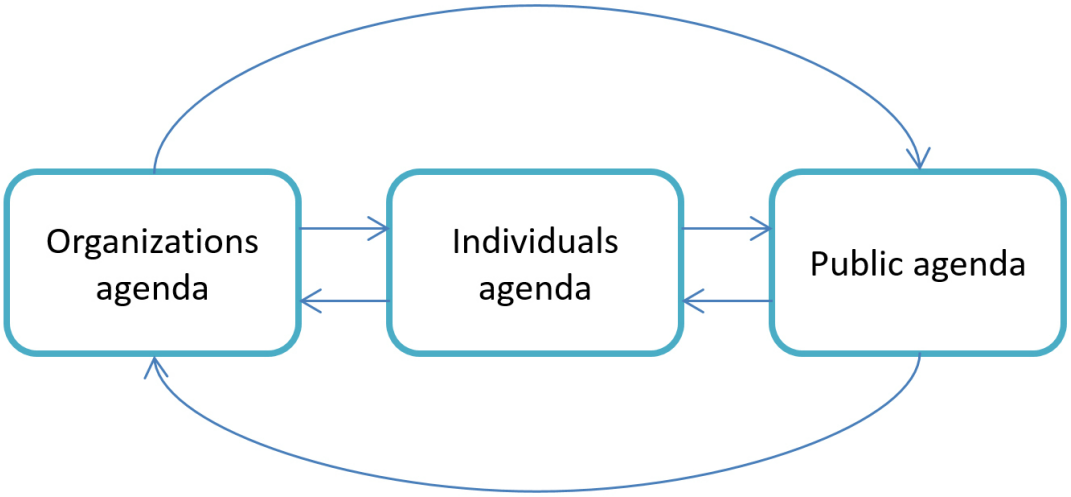
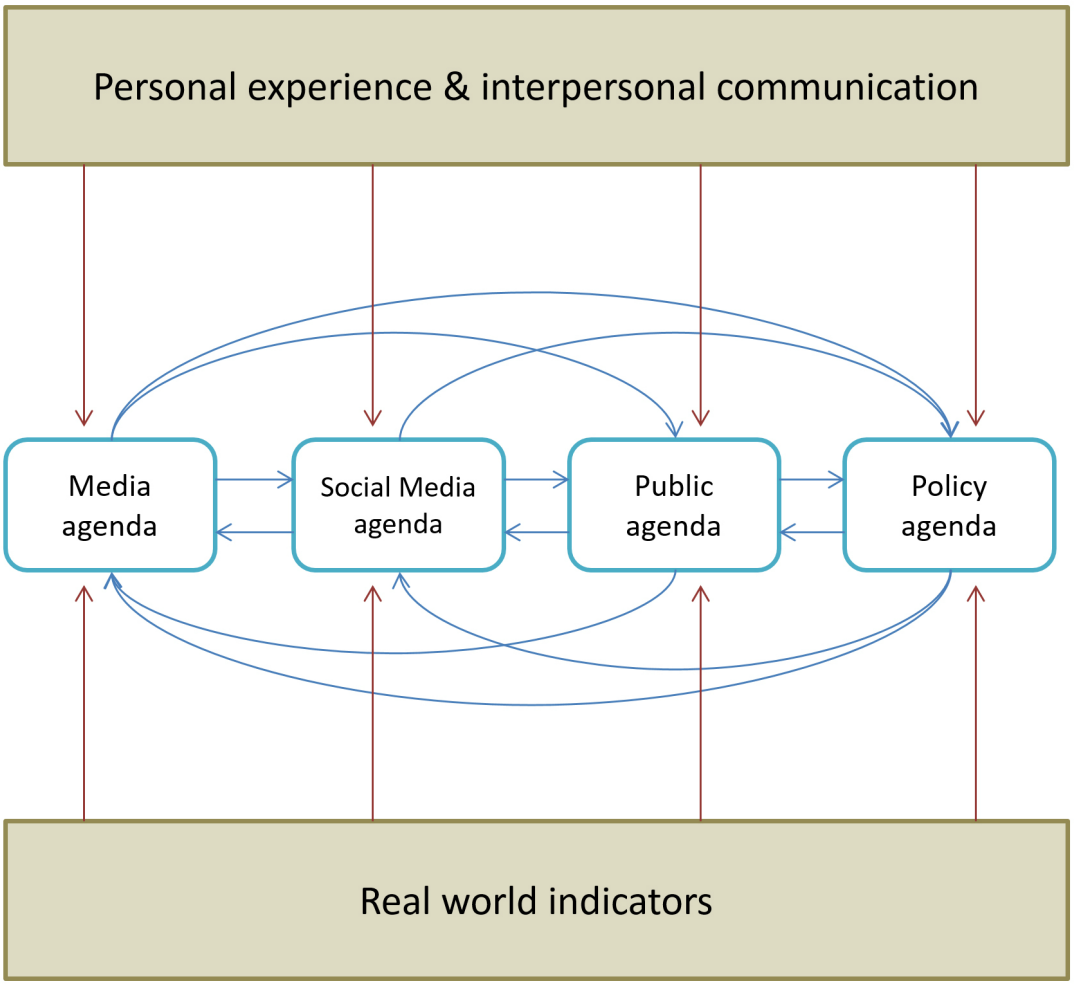


Figure 3. Proposed agenda setting model in the social media era.



Methods

Design

This study is part of exploratory research aimed at testing the capability of traditional communication theories in understanding social media platforms related to health promotion practice.

Exploratory research is preliminary research that contributes to the formulation and identification of hypotheses that show some merit in being followed up by confirmatory research [58,59]. The formulation of this hypothesis is not usually restricted, and a more flexible approach is used [59,60]. This study suggests appropriate hypotheses that fit with the aim of the study overall. In addition, the statistical tests used to examine these hypotheses, the analysis procedures employed, and the intervention that developed as part of the study harmonize the spirit of exploratory research [59,60].

The study used an important concern of public health, that of road traffic accidents. Based on the conceptual frames [61,62] and the selected message design [63,64], tweets about road traffic accidents were developed, pilot tested, and approved by a university research ethics committee. The study disseminated the tweets through the Saudi Ministry of Health Twitter account.

Immediately after completion of this Twitter intervention, a national campaign on road safety was run through various traditional and social media channels. The campaign enhanced the dynamics of mentioning the keywords of the study. Such enhancement does not bias the study as it reflects the normal dynamic of interactions targeted by the study to be examined. Furthermore, the collected data covered the periods before and after the campaign. Textbox 1 presents examples of intervention tweets as well as other users' tweets (see Multimedia Appendix 1 for the original Arabic texts).

Textbox 1. Examples of intervention tweets and tweets by other users that mention the keywords road traffic accident.

Intervention tweets:

- Good health is a major resource for social, economic, and personal development and an important dimension of quality of life.
- 81% of all deaths in Ministry of Health hospitals are due to road traffic accidents.
- Road safety is a public health issue which involves health as well as other sectors that have the responsibility to be engaged in road traffic accident prevention.

Tweets by other users:

- Your life is a candle; do not extinguish it. Stats: it is estimated that the number of traffic accidents will reach 1 million in the next 8 years.
- Traffic accidents cost SR 1.9 trillion (US $518 billion) globally each year.
- Did you know that in Saudi Arabia one person every hour is killed in a traffic accident?!!

Data Collection

Twitter data can be accessed directly from service profiles. Many third-party providers also offer Twitter statistics and analysis services. For instance, Tweetreach offers licensed access to the full Twitter "firehose" through Gnip, a licensed data reseller [65]. Increasing numbers of researchers are using these tools [66-72]. Account type was used as a variable for determining the most effective contributors to promoting road traffic accident agendas. Mentions of specific keywords as well as related variables (who tweeted messages, when messages were tweeted, to whom tweets were addressed, and how messages were tweeted) were the indicators used to measure contribution.

To validate the hypotheses, we collected data from two primary sources: Twitter activities and Saudi newspapers. As previously stated, keyword mentions served as indicators of agenda promotion; for Twitter, interactions were used to measure the process of agenda setting within the platform. Data from both sources were collected in a 90-day period from January 1 to March 31, 2014. Three Arabic keywords that are highly associated with road traffic accidents were considered in the analysis. The English translations of these keywords are "traffic accidents," "the traffic accidents," and "road accidents." Tracking mentions of predefined keywords have been used in previous studies although for different research purposes [71,73,74].

In collecting the Twitter data, we used the Tweetreach service to collect all tweets that contained the keywords. Many researchers have used this tool. For example, it has been employed in examining the use of Twitter as a platform for sharing information about medical events [75,76] and as a tracker and analysis tool in evaluating the effect of public health campaigns such as tobacco control social media advocacy [77].

In this study, we set up operators to filter tweets: only those expressed in the Arabic language were included, and tweets to and from Arab states other than Saudi Arabia were excluded. The final dataset comprised 59,046 tweets (16,071 regular tweets, 2783 replies, and 40,193 retweets) and 38,066 contributors. For each tweet and contributor, variables were collected. A total of three datasets were obtained from Twitter trackers: tweet data, retweet data, and contributor data.

In collecting the newspaper data, we used Google Advanced Search to gather information on six popular Saudi national newspapers: *Al-riyadh*, *Okaz*, *Al-Madina*, *Al-Yaum*, *Al-Watan*, and *Al-Jazirah*. Across these newspapers, 518 keyword mentions were recorded. All the datasets were extracted and prepared using Excel spreadsheets (Microsoft Corp).

Classification

The classification of data from social media, particularly Twitter, is coherent with the concept of infodemiology as described by Eysenbach [16,17] within which the study is framed/positioned.

For H1, we developed a classification to code the 2364 filtered users into four types of accounts. Individual accounts refer to any account owned by one person. Organizational accounts are those owned by a group or organization but not by media or policy organizations. Media accounts are accounts related to the media, including traditional media (programs or organizations) or news. Policy accounts are nonindividual accounts created for policy purposes or owned by political organizations.

Most of the accounts were classified in a straightforward manner as one of the authors is familiar with the Saudi environment; however, we needed to check profiles and tweets for some accounts. To validate the classification, an external observer independently classified 20% of the sample, made up of randomly chosen users from the list. The kappa [78] indicating interrater reliability was .87, indicating excellent agreement between our classification and that made by the external observer. The benchmark scale for strength of agreement proposed by Fleiss et al [79] was adopted in evaluating agreement between the study classification and that of the external observer (<40, poor; .40-.75, intermediate-good; >75, excellent).

Preparation and Analysis

Tweet Dataset

From the 59,046 Twitter activities, we obtained data on 16,073 regular tweets (not retweets or replies) that mentioned any of the three Arabic keywords. Each tweet was linked to user name, time, and variables listed in Table 1.

To isolate the influential tweets, tweets with no retweets were excluded (2895 tweets). From the dataset, we extracted the account users. A retweet impression indicates the ability of a user to reach audiences that extend beyond his/her direct followers. For users with more than one tweet we selected the tweet with the highest retweet impressions (1818 tweets). By manually checking the Twitter profiles of users, we filtered out all but user accounts owned by Saudi individuals or organizations with mainly Saudi audiences. Users on the list of 1115 were classified into the four account types.

The first step in the analysis was determining the degree of influence of the groups by calculating the total impressions for each classification type. Total impressions can be an informative measure of reach, which includes all the times at which a tweet was received (including receipt by the user's followers). Users with numerous followers can be influential in the Twitter community because of previously built influence. These users can be called Twitter influentials. To evaluate the ability of ordinary users to influence other users, we also analyzed retweet impressions, which show the total number of times a tweet was indirectly received. This is a strong measure of degree of influence, even among users with a limited number of followers. Subsequently, a simple rank scaling measure was applied to all the users on the list. We summed the total values of all the account types. For each of the three statistics (total impressions, total retweet impressions, and retweet impressions rank scaling), we calculated the percentage of the classification types and determined the average of the retweet impressions for these classifications. This multiple test technique allows more accurate assessment in examining the study hypotheses.

Retweet Dataset

From the primary Twitter dataset, we obtained a list of 38,066 contributors who mentioned the keywords during the 90-day data collection period. For each contributor, we used multiple variables, including the number of retweets by user, number of impressions, and amplification multipliers. Tweetreach defines impressions as the "number of timelines that received the tweet directly from the user" and the amplification multiplier as the "rate of amplification, based on how far that contributor's tweets spread due to the retweets and replies." The amplification rate is calculated as follows [80]:

[(total exposure − impressions) / impressions] + 1

Users with amplification multipliers below 1.2x were filtered out in accordance with the Tweetreach evaluation [81]: "anyone with an amplification multiplier of 1.2x or higher is doing quite well at spreading conversation." The final list included 1246 users who were coded in the classification stage. To analyze the data, we computed the total of the amplification multipliers calculated for the percentages of each classification type. We also calculated the average of the amplification multipliers for each type.

Contributors Dataset

Eysenbach [16], in relation to the concept of infodemiology, considers that advanced methods are required to explore the data from social networks and analyze the structures of interactions for public health. This study, rather than just identifying the presence of relationships between users on Twitter, interprets the data of contributor interactions to investigate the proposed hypotheses.

Based on total delivery ranks, we used the data on 40,193 retweets to extract data on 2665 retweets. These were all retweets over the average of deliveries, which was 2888.5. For each retweet, we identified users who retweeted a message and those who created the retweeted message. After filtering for both lists of users, 1951 unique users were classified. Type codes were used to identify 1382 users who interacted with one another.

Using R open source statistical computing and graphics software [82] we performed network analysis to explore the influence relationship among the four types of accounts. Statistics of edge interactions were calculated for each relationship, and a visualization graph was created.

Mention Trends

From the Twitter and newspaper data, we extracted two lists: total number of mentions in the examined newspapers and total number of mentions on Twitter. To normalize the data, all the

values were divided by the maximum value in each data column, after which the data were plotted on a simple line graph. Visually, similar spikes (ie, increases in mentions) were identified in the trend mentions of the newspapers and Twitter.

Results

Individual and Organizational Agenda

Among the 1115 users who posted regular tweets, the number of individual accounts was considerably higher than the number of other accounts. On the basis of the total of the three measures, individual users accounted for 76.29%, 67.79%, and 96.16% of retweet impressions, total impressions, and total amplification multipliers, respectively, as shown in Table 2.

Table 2. Percentage totals of the three measures for the four account types.

Account types	Retweet impressions rank scale (%)	Total impressions (%)	Amplification multipliers (%)	Average (%)
Individual	76.29	67.79	96.16	80.08
Organizational	11.94	15.23	0.02	9.06
Media	9.34	11.15	0.01	6.83
Policy	2.43	5.83	3.81	4.03

As determined from the three measures, the organizational accounts dominated the media and policy accounts in terms of retweet impressions and total impressions but not in total amplification multipliers. This result indicates that policy accounts are more influential than organizational accounts based on the amplification multiplier measure.

The average of the three totals (presented in Table 2) support H1-1a and H1-1b, which project individual accounts as the dominant type and organizational accounts as more influential than media and policy accounts.

The data on contributor interactions confirm the findings derived on the basis of the three measures (Table 3). The bidirectional individual accounts dominated. As a source of influence, the individual accounts registered 80.46% influence compared with the other account types; as a target, these accounts registered 59.33% influence. Organizational accounts (17.00% and 14.74% as source and target, respectively) also exhibited higher influence than did the media and policy accounts (1.45% and 1.09% as source, respectively, and 21.27% and 4.92% as target, respectively).

Table 3. Contributor interactions (based on retweeting relationship). For each account type, the table illustrates how much the tweet was retweeted by other account types.

Account type	Individual (n)	Organizational (n)	Media (n)	Policy (n)	Total (n)	Percentage (%)
Individual	672	125	267	48	1112	80.46
Organizational	132	68	18	17	235	17.00
Media	8	2	9	1	20	1.45
Policy	8	5	0	2	15	1.09
Total	820	200	294	68	1382	—
Percentage (%)	59.34	14.47	21.27	4.92	—	—

Individual Personal Agenda

In relation to the hypothesis on the power of individual accounts to influence the agendas of other Twitter users, the results do not revolve around the differentiation between individual and public agendas. Total statistical results are usually an informative indicator of mass influence, independent of the value added by the number of accounts to the total. To determine the influence of a single account, therefore, we calculated the average retweet impressions to examine the influence of both individual and public agendas. The averages indicate sustained influence of the individual agenda, which registered 72% influence in terms of total impressions and 96% in terms of the average amplification multiplier. On the basis of the contributor interactions, we assume that when an account is targeted by other account types, these accounts represent individual agendas rather than public agendas, especially when targeted individuals are not influentials or opinion leaders, as indicated by H1-2. As previously presented, the data show that the individual accounts registered 59.33% influence.

Influence of Agenda, Not Account

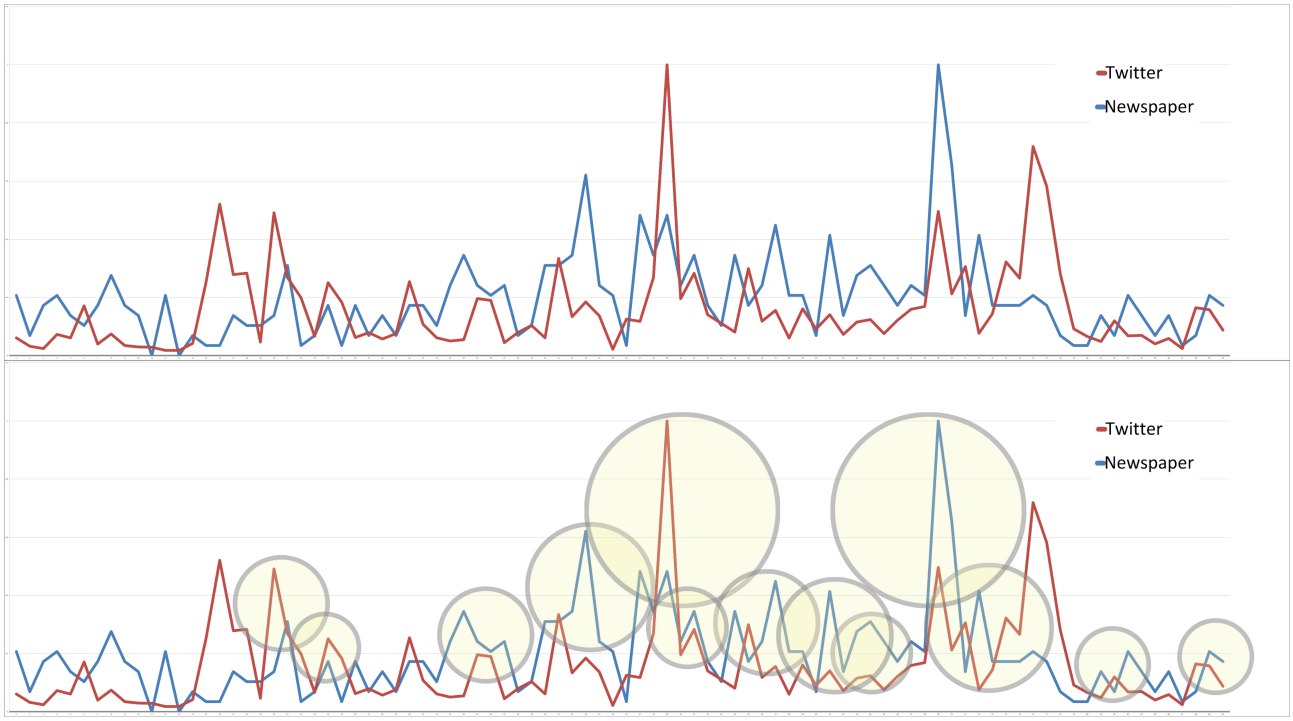
As proposed, a critical requirement here is to evaluate whether influence goes beyond accounts with special characteristics. The influence of agendas, rather than the influence of accounts, was validated by two correlations. The number of deliveries and retweet impressions exhibited a very weak correlation

(r=.08); the impressions and amplification multipliers showed a strong correlation (r=.01).

Twitter and Newspaper Agenda

The line graph that represents the mention trends in Twitter and newspapers (Figure 4) shows the relationship between these two time series datasets. Broadly, we identified 13 striking similarities between the two trends.

Figure 4. Line graph of the mention trends of Twitter and newspapers.



Discussion

Overview

Available evidence reflects the importance of communication technologies and the Internet in daily life and social interaction [43,46]. The characteristics and features of social media facilitate the powerful effects of such platforms in terms of disseminating information, framing opinions, and mobilizing action [43,46,83].

Among the many effective features of social media, user-generated content is critical to the positioning of new channels within the agenda-setting process. Other social media functions that influence agenda-setting dynamics are the sharing of content and the selection of the type of information that users want to receive [3].

Based on a simple classification scheme, Twitter accounts can be either personal or organizational.. This study suggests that in many cases, the emergence of agenda setting among social media users (eg, Twitter users) occurs through the advocacy of individual personal agendas. On social media platforms, individuals can influence the public and organizations through their own agenda. In other words, individuals can function as independent actors in agenda setting. In a similar vein, social media platforms enable organizations and the public to influence individual perceptions and behaviors through their agendas.

From the viewpoint of health promotion, agenda setting is an effective approach to achieving best practice aims and objectives [10,11,12]. Agenda setting can hold more potential than behavioral change strategies, as confirmed by road safety research and interventions [33,84,85]. This is demonstrated through the impact of influencing social policies [84,86], persuading policy and decision makers [7], orienting media coverage [11], enhancing the advocacy process, and maximizing the diffusion of innovations [12]. In addition to influencing health policy actions, the outcomes can result in positive changes in the behaviors of individuals [87]. This indirect approach is based on the view that human behaviors are not isolated from social and community contexts [38]. The evolution of the social media age encourages health promotion practitioners and organizations to maximize the benefits of innovations arising from such developments.

This exploratory study centers on the importance of a comprehensive understanding of best practices for health promotion. It does so by suggesting a novel contribution to health promotion through the development of an adapted agenda-setting approach in the social media era and within its platforms.

Agenda Setting in the Social Media Era

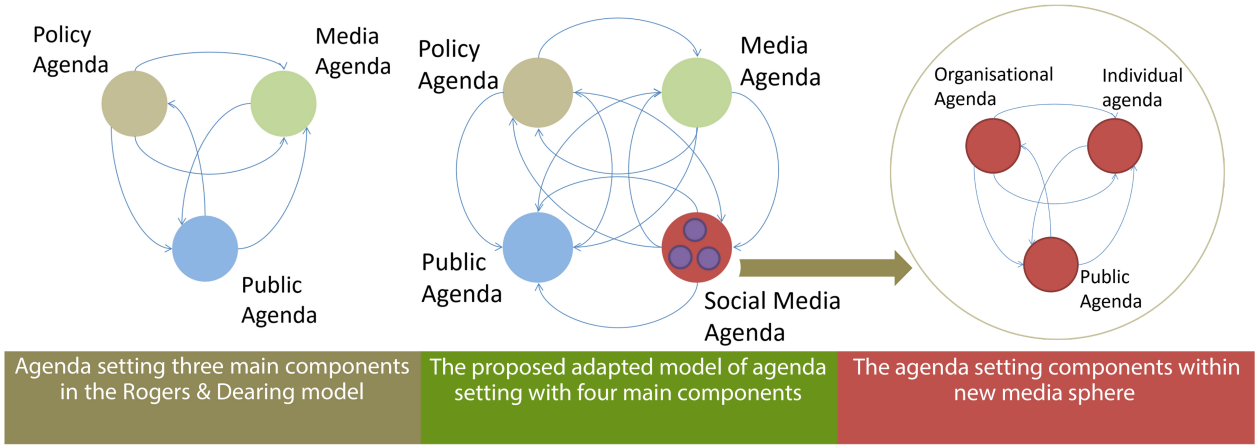
Health promotion needs a more creative approach to research and practice in using the agenda-setting function [33]. This is more challenging in the social media era where understanding

such theories requires more advanced research to keep pace with the evolution of this domain.

Unlike classic agenda setting, agenda setting as a process of interaction within the social media sphere involves different actors. With evolving tools and powerful features, social media provides a unique social space characterized by rich platforms where community members can communicate and interact. Agenda setting in the context of social media remains a communication process, but it differs from traditional communication [49] in that any individual or organization can be part of the social media communication dynamic. Agenda setting via social media is a new sphere of social interaction that represents any member who desires to participate in the process. Moreover, the growth of social media use among communities enables this sphere to influence daily life.

Given this backdrop, we suggest two levels of understanding in exploring agenda setting in the social media era. First, we propose social media as an independent agenda-formulating body within the agenda-setting process. Second, we recommend the exploration of agenda setting within social media. Figure 5 illustrates the development stages of these proposed adapted models.

Figure 5. The development stages of the study proposed models built on two levels of understanding of the agenda setting process.



Individual and Organizational Agendas

We argue that within Twitter, agenda-setting participants can change, unlike the fixed nature of participants in classic agenda setting. This variability is attributed to the power shift towards new actors, such as individuals and organizations, as indicated in previous research [46,48,49]. We hypothesized that individuals serve as new actors in the process and that they can formulate their own personal agendas instead of adhering only to the public agenda. In addition, media and policy agendas, as part of organizational agendas, may extend to the agendas of different organizations.

The results suggest that the individual agenda dominated over the other agenda types and that organizations exhibited stronger influence than that wielded by media and policy groups. Although the data of amplification multiplier measure showed that policy accounts are more influential than organizational accounts, this influence is limited to retweets of other accounts only, which means policy accounts hold more ability to enhance the diffusion of tweets by influencing nonfollower users. In spite of this, each of the three measures used to evaluate the influence of the account types presented analogous results from various calculation methods and different variables. We believe this feature strengthens the evidence supporting the formulated hypotheses because it rectifies the limitations of one indicator or its measures.

Individual Personal Agenda

For the individual agenda, the analysis derived different statistics which confirm that power is not restricted to the public agenda but extends to individual personal agendas. The averages of the total impressions and amplification multipliers suggest the influence of the personal agenda. The contributor interactions also support this finding, as indicated by the strong influence of the individual agenda as a target. The findings of the data analyses also highlight the influence of the public agenda, but the statistical results do not demonstrate specific differences between personal and public agendas. Agenda melding among individuals was reflected by the data on contributor interactions (672 interactions recorded).

Influence of Agenda, Not Account

This study aimed to examine the effectiveness of agenda setting as a theory in the social media era. Therefore, a critical requirement was to consider the influence of agenda with reference to the type of account rather than account influence. The dynamics of interactions on Twitter are affected by many factors and will be/are reflected in the data [88]. A highly influential account would have generated strong bias in such an examination if it affected the values calculated on the basis of the study data. The correlation results confirm that the popularity of accounts was not an issue in the derived data values.

Influentials Significance in Agenda-Setting Process

Although influentials or opinion leaders are beyond the scope of this work, they remain essential participants in any social communication process, including agenda setting [89]. In early theories such as the two-step flow [90] and diffusion of innovations [89], influentials have a significant impact on influencing the agenda and enhancing its diffusion. It is suggested by Kozel et al [33] that it is crucial for health promotion agenda-setting practices to develop strategies that enhance the diffusion of health promotion agendas through all agenda-setting process components. In addition to amplifying the diffusion of influence, these influentials can play a key role in sustaining the salience of specific issues, which contributes to effective agenda-setting processes [33]. The data yielded by the measures used in this study show the high impact of influentials on the agenda among users, which will increase the interrelation dynamics of influencing the agenda among different account types. Thus, using influentials is one of the strategies that can be efficient in health promotion agenda-setting practices on social media platforms. The role of influentials requires further research which is beyond the scope of the current study.

Twitter and Newspaper Agenda

We hypothesized a relationship between newspapers and Twitter and illustrated this association via a line graph of mention trends over the platform. Nevertheless, the collected data cannot illuminate a clear direction for this relationship and do not exclude external factors that can affect the trends. This hypothesis was intended as a starting point in exploring the incorporation of the social media agenda in the classic agenda-setting process. This relationship has been studied and validated through various analysis techniques and methodologies [44,45,48,49,55,91]. Furthermore, examining causality is a crucial component of determining the direction of the relationship between new and traditional media in agenda setting [48,92].

Nonetheless, this relationship is only part of the model proposed in this work and therefore requires further examination [44,83]. Particularly interesting focal issues in this regard are the agenda-setting interaction between social media and the public agenda and that between social media and the policy agenda [44].

Limitations

We have proposed an alternative perspective from which to understand agenda setting in the social media age. Supported by the data collected, we adopted well-defined measures that reflected a positive evaluation of the study's hypotheses. We have applied multiple hypothesis testing methods which can support authenticity of the study as is recommended for any exploratory research [59]. Nonetheless, it is crucial in scientific research that any exploratory research must be carefully examined by confirmatory research [59].

The nature of data on social media results in limitations which can affect any infodemiology study. These include, for example, the lack of strong representation of the population and the lack of accuracy of some information provided on such platforms [17].

Furthermore, the statistical procedures and measures used require replication and/or repetition to establish a solid foundation for the implications of the results. As indicated by McCombs [45], such studies are limited by variables related to time, place, and the selection of measurement and analysis tools; a repeated examination of a model and the replication of ideas are thus critical requirements in the validation of results. Utilizing agenda setting in the field of health promotion will encounter many challenges in research and practice [93].

Many factors and variables must be considered, guided by a more in-depth understanding of the process of agenda setting in relation to its possible applications [93]. This will provide valuable strategies and themes for the successful application of agenda setting in promoting the public's health [93].

For the public health and health promotion domain, further research may include different health topics with larger data samples and modified methods. At the organizational level, the findings suggest more dominance for health organizations in the agenda-setting process within social media. Further research related to strategies and best practices is required for such organizations to close the gap that has already been identified [53]. Moreover, such research may consider extending the infodemiology framework to development efforts addressing more topics, languages, and platforms.

Conclusion

The results indicate that media platforms are a promising avenue that can enhance the efficacy of intervention programs. However, the effective use of such platforms will necessitate new strategies that address the limitations of traditional communication channels. More efforts towards modifying health promotion strategies and developing new approaches should be initiated. For such development, conducting research is vital to establishing a strong basis for the design, formulation, and implementation of agendas. Social media augments the effectiveness of this approach by shifting power towards reachable individual participants. In turn, agendas become more accessible and more easily used as tools or targets of influence. Finally, organizations that promote the public's health will benefit considerably from actively participating in the agenda-setting process through the formulation of the organizational agenda.

Acknowledgments

We would like to acknowledge Yahya Albalawi's advice and guidance in relation to the statistical analysis.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Examples of intervention tweets and tweets by other users that mention the keywords road traffic accident (Arabic texts).

[PDF File (Adobe PDF File), 295KB - publichealth_v1i2e21_app1.pdf]

References

1. Frati A, Luzi A, Colucci A. Communication for health promotion: history and identification of effective methods. Ann Ist Super Sanità 2010;46(4):422-426 [FREE Full text] [doi: 10.4415/ANN_10_04_10]
2. Suresh K. Evidence based communication for health promotion: Indian lessons of last decade. Indian J Public Health 2011;55(4):276-285 [FREE Full text] [doi: 10.4103/0019-557X.92405] [Medline: 22298137]
3. 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]
4. World Health Organization. Geneva: World Health Organization (WHO); 2013. Global status report on road safety Internet URL: http://www.who.int/iris/bitstream/10665/78256/1/9789241564564_eng.pdf?ua=1 [WebCite Cache ID 6acr0b6Be]
5. Ageli M, Zaidan A. Road Traffic Accidents in Saudi Arabia: An ADRL Approach and Multivariate Granger Causality. International Journal of Economics and Finance 2013;5(7). [doi: 10.5539/ijef.v5n7p26]
6. Ansari S, Akhdar F, Mandoorah M, Moutaery K. Causes and effects of road traffic accidents in Saudi Arabia. Public Health 2000 Jan;114(1):37-39. [doi: 10.1038/sj.ph.1900610] [Medline: 10787024]
7. Peden M., Scurfield R., Sleet D., Mohan D., Hyder A., Jarawan E., Mathers C.. World Health Organization. Geneva: World Health Organization; 2004. World report on road traffic injury prevention Internet URL: http://whqlibdoc.who.int/publications/2004/9241562609.pdf [WebCite Cache ID 6bfTSuJZy]
8. Balatsoukas P, Kennedy C, Buchan I, Powell J, Ainsworth J. J Med Internet Res(6). 2015. The Role of Social Network Technologies in Online Health Promotion: A Narrative Review of Theoretical and Empirical Factors Influencing Intervention Effectiveness URL: http://www.jmir.org/2015/6/e141/ [WebCite Cache ID 6bfSoz9Lv]
9. Narayanaswami P, Gronseth G, Dubinsky R, Penfold-Murray R, Cox J, Bever JC, et al. J Med Internet Res(8). 2015. The Impact of Social Media on Dissemination and Implementation of Clinical Practice Guidelines: A Longitudinal Observational Study URL: http://www.jmir.org/2015/8/e193/ [WebCite Cache ID 6bfTwU9ze]
10. Leask J, Hooker C, King C. Media coverage of health issues and how to work more effectively with journalists: a qualitative study. BMC Public Health 2010;10:535 [FREE Full text] [doi: 10.1186/1471-2458-10-535] [Medline: 20822552]
11. Cohen L, Cha´vez V, Chehimi S. Prevention is primary: strategies for community well being. San Francisco, CA: Jossey-Bass; 2010.
12. Kozel CT, Kane WM, Rogers EM, Brandon JE, Hatcher MT, Hammes MJ, et al. Exploring health promotion agenda-setting in New Mexico: reshaping health promotion leadership. Promot Educ 2003;10(4):171-7, 198, 209. [Medline: 15071972]
13. Kaplan A, Haenlein M. Users of the world, unite! The challenges and opportunities of Social Media. Business Horizons 2010 Jan;53(1):59-68. [doi: 10.1016/j.bushor.2009.09.003]
14. Internet World Stats. World Internet Users Statistics and 2014 World Population Stats URL: http://www.internetworldstats.com/stats.htm [WebCite Cache ID 6aOgE6a6n]
15. Eysenbach G. Infodemiology: The epidemiology of (mis)information. Am J Med 2002 Dec 15;113(9):763-765. [Medline: 12517369]
16. 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]
17. Eysenbach G. Infodemiology and Infoveillance. American Journal of Preventive Medicine 2011;40(5):S154-S158. [doi: 10.1016/j.amepre.2011.02.006]
18. Eysenbach G. Infodemiology: Tracking Flu-Related Searches on the Web for Syndromic Surveillance. AMIA Annu Symp Proc 2006:244-248.
19. Zwolenski M, Weatherill L. The Digital Universe Rich Data and the Increasing Value of the Internet of Things. ajTDE 2014 Oct 20;2(3) [FREE Full text] [doi: 10.7790/ajtde.v2n3.47]
20. Bail C. The cultural environment: measuring culture with big data. Theor Soc 2014 Mar 27;43(3-4):465-482. [doi: 10.1007/s11186-014-9216-5]
21. Twitter Help Center. Getting started with Twitter URL: https://support.twitter.com/groups/50-welcome-to-twitter/topics/204-the-basics/articles/215585-getting-started-with-twitter [WebCite Cache ID 6aOlw1c4l]
22. Jason M. Global Web Index. London: Global Web Index; 2015. Twitter Users - Q1 2015 Internet URL: https://app.globalwebindex.net/products/report/twitter-users-q1-2015 [WebCite Cache ID 6aOm8d8vX]

23. Internet world stats.: Internet world stats Middle East Internet Usage Statistics, Population, Facebook and Telecommunications Reports Internet URL: http://www.internetworldstats.com/stats5.htm [WebCite Cache ID 6aOmX7iuT]
24. Lippmann W. Public opinion. New York: Macmillan; 1922.
25. Lasswell H. The structure and function of communication in society. In: Bryson L, editor. The communication of ideas: a series of addresses. New York: Institute for Religious and Social Studies; 1948:215-228.
26. Cohen B. Press and Foreign Policy. New Jersey: Princeton University Press; 1963.
27. McCombs M, Shaw D. The Agenda-Setting Function of Mass Media. Public Opinion Quarterly 1972;36(2):176-187. [doi: 10.1086/267990]
28. University of Utwente. 2015. Agenda Setting Theory URL: http://www.utwente.nl/cw/theorieenoverzicht/Theory%20clusters/Mass%20Media/Agenda-Setting_Theory/ [WebCite Cache ID 6aOnUb56H]
29. Rogers E, Dearing J. Agenda-setting research: Where has it been? Where is it going? In: Anderson J, editor. Communication yearbook. 11th ed. Newbury Park, CA: Sage; 1988:555-594.
30. Kosicki G. Problems and Opportunities in Agenda-Setting Research. J Communication 1993 Jun;43(2):100-127. [doi: 10.1111/j.1460-2466.1993.tb01265.x]
31. Berger B. Private Issues and Public Policy: Locating the Corporate Agenda in Agenda-Setting Theory. Journal of Public Relations Research 2009 Nov 19;13(2):91-126. [doi: 10.1207/s1532754xjprr1302_1]
32. Kozel C, Kane W, Rogers E, Hammes M. Exploring health promotion agenda-setting design: A study of the factors associated with health promotion agenda-setting design in New Mexico. Unpublished doctoral dissertation 1995 University of New Mexico, Albuquerque.
33. Kozel C, Kane W, Hatcher M, Hubbell A, Dearing J, Forster-Cox S. Introducing health promotion agenda-setting for health education practitioners. California Journal of Health Promotion 2006;4(1):32-40.
34. Kozel C, Hubbell A, Dearing J, Kane W, Thompson S, Perez F. Exploring agenda-setting for Healthy Border 2010: Research directions and methods. Californian Journal of Health Promotion 2006;4(1):141-161.
35. Farmer J, Kozel C. Agenda Setting and the Albuquerque Clean Indoor Air Ordinance. Journal of Health Education 2005;36(5):313-315.
36. Pratt C, Ha L, Pratt C. Setting the Public Health Agenda on Major Diseases in Sub-Saharan Africa: African Popular Magazines and Medical Journals, 1981--1997. Journal of communication. - 2002;52(4):889-904. [doi: 10.1093/joc/52.4.889]
37. Grover L, Horent E. Public health agenda setting in Louisiana: an applied index and implications for community development policy. Community Development 2011 Apr;42(2):268-285. [doi: 10.1080/15575330.2011.558207]
38. Terris M. A social policy for health. American Journal of Public Health and the Nations Health 1968;58(1):5-12.
39. Hao-yu F. An Interpretation of Agenda-Setting of the Health Communication on the Internet. Journal of Fujian Normal University (Philosophy and Social Sciences Edition) 2013;1:031.
40. Flora JA, Maibach EW, Maccoby N. The role of media across four levels of health promotion intervention. Annu Rev Public Health 1989;10:181-201. [doi: 10.1146/annurev.pu.10.050189.001145] [Medline: 2655630]
41. Molleman Gerard R M, Bouwens Jan G M. Building the evidence base: from tool development to agenda-setting and defining a joint programme for health promotion in Europe. Promot Educ 2005;Suppl 1:8-9, 43. [Medline: 15952272]
42. Pacheco J, Boushey G. Public health and agenda setting: determinants of state attention to tobacco and vaccines. J Health Polit Policy Law 2014 Jun;39(3):565-589. [doi: 10.1215/03616878-2682612] [Medline: 24603088]
43. Liang B, Scammon D. Incidence of Online Health Information Search: A Useful Proxy for Public Health Risk Perception. J Med Internet Res 2013;15(6):e114 [FREE Full text] [doi: 10.2196/jmir.2401]
44. Zheluk A, Gillespie J, Quinn C. Searching for Truth: Internet Search Patterns as a Method of Investigating Online Responses to a Russian Illicit Drug Policy Debate. J Med Internet Res 2012;14(6):e165 [FREE Full text] [doi: 10.2196/jmir.2270]
45. Dearing J, Rogers E. Agenda-setting. Thousand Oaks, Calif: Sage; 1996.
46. Bekkers V, Beunders H, Edwards A, Moody R. New Media, Micromobilization, and Political Agenda Setting: Crossover Effects in Political Mobilization and Media Usage. The Information Society 2011 Jul;27(4):209-219. [doi: 10.1080/01972243.2011.583812]
47. McCombs M. A Look at Agenda-setting: past, present and future. Journalism Studies 2005 Nov;6(4):543-557. [doi: 10.1080/14616700500250438]
48. Sayre B, Bode L, Shah D, Wilcox D, Shah C. Agenda Setting in a Digital Age: Tracking Attention to California Proposition 8 in Social Media, Online News and Conventional News. Policy & Internet 2010;2(2):7-32. [doi: 10.2202/1944-2866.1040]
49. Tran H. Online agenda setting: A new frontier for theory development. In: Thomas J. J, editor. Agenda Setting in a 2.0 World: New Agendas in Communication: A Tribute to Maxwell McCombs;-. Oxon: Routledge; 2014:205-229.
50. Althaus S, Tewksbury D. Agenda Setting and the "New" News: Patterns of Issue Importance Among Readers of the Paper and Online Versions of the New York Times. Communication Research 2002 Apr 01;29(2):180-207. [doi: 10.1177/0093650202029002004]
51. Conway M, Patterson J. Today's Top Story? An Agenda-Setting and Recall Experiment Involving Television and Internet News. Southwestern Mass Communication Journal 2008;24(1):31-48.

52. Neiger BL, Thackeray R, Burton SH, Giraud-Carrier CG, Fagen MC. Evaluating social media's capacity to develop engaged audiences in health promotion settings: use of Twitter metrics as a case study. Health Promot Pract 2013 Mar;14(2):157-162. [doi: 10.1177/1524839912469378] [Medline: 23271716]
53. 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]
54. Gulyas A. THE INFLUENCE OF PROFESSIONAL VARIABLES ON JOURNALISTS'USES AND VIEWS OF SOCIAL MEDIA: A comparative study of Finland, Germany, Sweden and the United Kingdom. Digital Journalism 2013 Jun;1(2):270-285. [doi: 10.1080/21670811.2012.744559]
55. Jones R, Rogers R, Roberts J, Callaghan L, Lindsey L, Campbell J, et al. What Is eHealth (5): A Research Agenda for eHealth Through Stakeholder Consultation and Policy Context Review. J Med Internet Res 2005;7(5):e54 [FREE Full text] [doi: 10.2196/jmir.7.5.e54]
56. Meraz S. Is There an Elite Hold? Traditional Media to Social Media Agenda Setting Influence in Blog Networks. Journal of Computer-Mediated Communication 2009;14(3):682-707. [doi: 10.1111/j.1083-6101.2009.01458.x]
57. Zhu J, Watt J, Snyder L, Yan J, Jiang Y. Public Issue Priority Formation: Media Agenda-Setting and Social Interaction. J Communication 1993 Mar;43(1):8-29. [doi: 10.1111/j.1460-2466.1993.tb01246.x]
58. Tukey J. We Need Both Exploratory and Confirmatory. The American Statistician 1980 Feb;34(1):23-25. [doi: 10.2307/2682991]
59. Goeman J, Solari A. Multiple Testing for Exploratory Research. Statist. Sci 2011 Nov;26(4):584-597. [doi: 10.1214/11-sts356]
60. Koestler Parapsychology Unit Study Registry. Edinburgh: University of Edinburgh; 2015. Exploratory and Confirmatory Analyses Internet URL: http://www.koestler-parapsychology.psy.ed.ac.uk/Documents/explore_confirm.pdf [accessed 2015-09-20] [WebCite Cache ID 6bgqIZdXw]
61. Lakoff G, Morgan P. Framing social issues: Does "the working poor" work? In: Report to the Ford Foundation; Rockridge Strategic Analysis Paper 2001-01. Cambridge, MA: Kennedy School of Government, Harvard University; 2001 Presented at: Public obligations: Giving kids a chance; June 1, 2001; Cambridge p. 16-38.
62. Dorfman L, Wallack L, Woodruff K. More than a message: framing public health advocacy to change corporate practices. Health Educ Behav 2005 Jun;32(3):320-36; discussion 355. [doi: 10.1177/1090198105275046] [Medline: 15851542]
63. O'Flahavan L, Goulet A. Centers for Disease Control and Prevention. 2012. CDC's Guide to Writing for Social Media URL: http://www.cdc.gov/socialmedia/Tools/guidelines/pdf/GuidetoWritingforSocialMedia.pdf [WebCite Cache ID 6bgrZUNIE]
64. Public Health Ontario; Dalla Lana School of Public Health. Toronto, ON: Queen's Printer for Ontario; 2012. Health communication message review criteria URL: http://www.publichealthontario.ca/en/eRepository/Health_communication_message_review_criteria_2012.pdf [accessed 2015-09-20] [WebCite Cache ID 6bgrhTGZB]
65. Union Metrics.: Union Metrics; 2014. TweetReach: Where our Twitter data comes from - TweetReach Blog Internet URL: https://blog.tweetreach.com/2014/01/tweetreach-where-our-twitter-data-comes-from/ [accessed 2015-07-30] [WebCite Cache ID 6aOnHlUVd]
66. Anger I, Kittl C. Measuring influence on Twitter. New York: ACM; 2011 Presented at: the 11th International Conference on Knowledge Management and Knowledge Technologies (i-KNOW '11); September 07 - 09, 2011; Graz, Austria p. 31:1-31:4. [doi: 10.1145/2024288.2024326]
67. Bakshy E, Hofman J, Mason W, Watts D. Everyone's an influencer: quantifying influence on twitter. New York, NY: ACM; 2011 Presented at: Fourth ACM International Conference on Web Search and Data Mining (WSDM '11); February 9-12, 2011; Hong Kong, China p. 65-74. [doi: 10.1145/1935826.1935845]
68. Kwak H, Lee C, Park H, Moon S. What is Twitter, a social network or a news media? New York: ACM; 2010 Presented at: The 19th international conference on World wide web (WWW '10); 26-30 April 2010; Raleigh, North Carolina, USA p. 591-600. [doi: 10.1145/1772690.1772751]
69. Cha M, Haddadi H, Benevenuto F, Gummadi P. Measuring User Influence in Twitter: The Million Follower Fallacy. 2010 Presented at: ICWSM. (10-17); 23-26 May 2010; Washington, DC p. 30.
70. Wu S, Hofman J, Mason W, Watts D. Who says what to whom on twitter. New York: ACM; 2011 Presented at: The 20th international conference on World wide web (WWW '11); 28 March- 1 April 2011; Hyderabad, India p. 705-714. [doi: 10.1145/1963405.1963504]
71. Mollema L, Harmsen I, Broekhuizen E, Clijnk R, De MH, Paulussen T, et al. Disease Detection or Public Opinion Reflection? Content Analysis of Tweets, Other Social Media, and Online Newspapers During the Measles Outbreak in the Netherlands in 2013. J Med Internet Res 2015;17(5):e128 [FREE Full text] [doi: 10.2196/jmir.3863]
72. Kendra R, Karki S, Eickholt J, Gandy L. Characterizing the Discussion of Antibiotics in the Twittersphere: What is the Bigger Picture? J Med Internet Res 2015;17(6):e154 [FREE Full text] [doi: 10.2196/jmir.4220]
73. McIver D, Hawkins J, Chunara R, Chatterjee A, Bhandari A, Fitzgerald T, et al. Characterizing Sleep Issues Using Twitter. J Med Internet Res 2015;17(6):e140 [FREE Full text] [doi: 10.2196/jmir.4476]
74. Weeg C, Schwartz H, Hill S, Merchant R, Arango C, Ungar L. Using Twitter to Measure Public Discussion of Diseases: A Case Study. JMIR Public Health Surveill 2015 Jun 26;1(1):e6 [FREE Full text] [doi: 10.2196/publichealth.3953]

75. Cochran A, Kao L, Gusani N, Suliburk J, Nwomeh B. Use of Twitter to document the 2013 Academic Surgical Congress. Journal of Surgical Research 2014;190(1):36-40. [doi: 10.1016/j.jss.2014.02.029]
76. Udovicich C, Barberi A, Perera K. Tweeting the meeting: a comparative analysis of an Australian emergency medicine conference over four years. J Emerg Trauma Shock 2015 [FREE Full text] [doi: 10.4103/0974-2700.161655]
77. Hefler M, Freeman B, Chapman S. Tobacco control advocacy in the age of social media: using Facebook, Twitter and change. Tob Control 2013 May;22(3):210-214. [doi: 10.1136/tobaccocontrol-2012-050721] [Medline: 23047890]
78. Cohen J. A Coefficient of Agreement for Nominal Scales. Educational and Psychological Measurement 1960 Apr 01;20(1):37-46. [doi: 10.1177/001316446002000104]
79. Fleiss L, Bruce L, Myunghee C. The measurement of interrater agreement. In: Statistical Methods for Rates and Proportions. 3rd ed. Hoboken, NJ, USA: John Wiley & Sons Inc; 1981:212-236.
80. Betsy Rowden. Union Metrics Help Desk.: Union Metrics; 2015. Understanding The Tracker CSV Export Internet URL: http://help.tweetreach.com/hc/en-us/articles/202899469-Understanding-The-Tracker-CSV-Export [accessed 2015-07-30] [WebCite Cache ID 6aOoB3GTu]
81. Union Metrics, TweetReach Blog Internet.: Union Metrics Measuring participant influence through amplification on Twitter URL: https://blog.tweetreach.com/2011/11/measuring-participant-influence-through-amplification-on-twitter/ [accessed 2015-07-30] [WebCite Cache ID 6aOpOi61Z]
82. The R Project for Statistical Computing. 2015. URL: http://www.r-project.org/ [WebCite Cache ID 6aOpK5Aso]
83. Roberts M, Wanta W, Dzwo T. Agenda Setting and Issue Salience Online. Communication Research 2002 Aug 01;29(4):452-465 [FREE Full text] [doi: 10.1177/0093650202029004004]
84. Elder RW, Shults RA, Sleet DA, Nichols JL, Thompson RS, Rajab W, Task Force on Community Preventive Services. Effectiveness of mass media campaigns for reducing drinking and driving and alcohol-involved crashes: a systematic review. Am J Prev Med 2004 Jul;27(1):57-65. [doi: 10.1016/j.amepre.2004.03.002] [Medline: 15212776]
85. Heng Kenneth W J, Vasu A. Newspaper media reporting of motor vehicle crashes in Singapore: an opportunity lost for injury prevention education? Eur J Emerg Med 2010 Jun;17(3):173-176. [doi: 10.1097/MEJ.0b013e328330b40a] [Medline: 19704376]
86. Farmer J, Kozel C. Agenda Setting and the Albuquerque Clean Indoor Air Ordinance. American Journal of Health Education 2005 Oct;36(5):313-315. [doi: 10.1080/19325037.2005.10608202]
87. Moon S. Attention, Attitude, and Behavior: Second-Level Agenda-Setting Effects as a Mediator of Media Use and Political Participation. Communication Research 2011 Oct 17;40(5):698-719. [doi: 10.1177/0093650211423021]
88. Mustafaraj E, Finn S, Whitlock C, Metaxas P. Vocal minority versus silent majority: Discovering the opionions of the long tail. : IEEE; 2011 Presented at: In: Privacy, Security, Risk and Trust (PASSAT) and IEEE Third International Conference on Social Computing (SocialCom); 9-11 Oct. 2011; Athens, Greece p. 103-110. [doi: 10.1109/PASSAT/SocialCom.2011.188]
89. Rogers E. Diffusion of innovations. New York: Free Press; 1983.
90. Lazarsfeld P, Berelson B, Gaudet H. The peoples choice: how the voter makes up his mind in a presidential campaign. In: Psychological Bulletin. New York: Columbia University Press; 1948.
91. Broersma M, Graham T. TWITTER AS A NEWS SOURCE: How Dutch and British newspapers used tweets in their news coverage, 2007–2011. Journalism Practice 2013 Aug;7(4):446-464. [doi: 10.1080/17512786.2013.802481]
92. Tan Y, Weaver D. Agenda-Setting Effects among the Media, the Public, and Congress, 1946-2004. Journalism & Mass Communication Quarterly 2007 Dec 01;84(4):729-744. [doi: 10.1177/107769900708400405]
93. Moran C, Kozel C, Hubbell A. Turning Research into Practice: Key Strategies for Developing a Shared Vision Approach for Health Education Advocacy. Californian Journal of Health Promotion 2011;9(2):77-84.

Edited by G Eysenbach; submitted 10.08.15; peer-reviewed by C Kozel, S Bhattacharya; comments to author 02.09.15; revised version received 20.09.15; accepted 17.10.15; published 25.11.15.

Please cite as:
Albalawi Y, Sixsmith J
Agenda Setting for Health Promotion: Exploring an Adapted Model for the Social Media Era
JMIR Public Health Surveill 2015;1(2):e21
URL: http://publichealth.jmir.org/2015/2/e21/
doi:10.2196/publichealth.5014
PMID:27227139

Original Paper

Identifying Adverse Effects of HIV Drug Treatment and Associated Sentiments Using Twitter

Cosme Adrover[1], PhD; Todd Bodnar[1], BSc; Zhuojie Huang[1], PhD; Amalio Telenti[2*], MD, PhD; Marcel Salathé[1*], PhD

[1]Center for Infectious Disease Dynamics, Department of Biology, Penn State University, University Park, PA, United States

[2]J. Craig Venter Institute, La Jolla, CA, United States

[*]these authors contributed equally

Corresponding Author:
Marcel Salathé, PhD
Center for Infectious Disease Dynamics
Department of Biology
Penn State University
MSC W-251
University Park, PA, 16803
United States
Phone: 1 4083868916
Fax: 1 814 863 7590
Email: salathe.marcel@gmail.com

Abstract

Background: Social media platforms are increasingly seen as a source of data on a wide range of health issues. Twitter is of particular interest for public health surveillance because of its public nature. However, the very public nature of social media platforms such as Twitter may act as a barrier to public health surveillance, as people may be reluctant to publicly disclose information about their health. This is of particular concern in the context of diseases that are associated with a certain degree of stigma, such as HIV/AIDS.

Objective: The objective of the study is to assess whether adverse effects of HIV drug treatment and associated sentiments can be determined using publicly available data from social media.

Methods: We describe a combined approach of machine learning and crowdsourced human assessment to identify adverse effects of HIV drug treatment solely on individual reports posted publicly on Twitter. Starting from a large dataset of 40 million tweets collected over three years, we identify a very small subset (1642; 0.004%) of individual reports describing personal experiences with HIV drug treatment.

Results: Despite the small size of the extracted final dataset, the summary representation of adverse effects attributed to specific drugs, or drug combinations, accurately captures well-recognized toxicities. In addition, the data allowed us to discriminate across specific drug compounds, to identify preferred drugs over time, and to capture novel events such as the availability of preexposure prophylaxis.

Conclusions: The effect of limited data sharing due to the public nature of the data can be partially offset by the large number of people sharing data in the first place, an observation that may play a key role in digital epidemiology in general.

(JMIR Public Health Surveill 2015;1(2):e7) doi:10.2196/publichealth.4488

KEYWORDS

Twitter; HIV; AIDS; pharmacovigilance; mTurk; mechanical Turk

Introduction

The Sharing of Health Information on Twitter

Twitter is a popular microblogging platform where users publicly share information, including personal thoughts and emotions. Everyday, hundreds of millions of tweets are posted on Twitter. This offers a large potential source of information for public health purposes. The sharing of information about personal health is now widely recognized to be a broad phenomenon that occurs in almost any field imaginable [1].

Examples include the sharing of vaccination behavior [2], the sharing of marijuana consumption [3], the sharing of weight loss attempts [4], or the sharing of suicidal thoughts [5]. Moreover, this large pool of available data can be used to estimate the most common side effects of certain pharmaceutical products [6,7]. In this study, we will focus on identifying human immunodeficiency virus (HIV)-infected individuals currently undergoing drug treatment, and on their personal experiences, specifically, with regard to drug toxicities.

We chose a specific question that a priori seemed difficult: to identify infected individuals that would communicate about their HIV status, and more specifically, about their treatment. Drug adverse effects have been a source of concern, both for the medical establishment, and for HIV-infected individuals, due to their prevalence [8,9], and because treatments have to be taken for life. There is a general agreement that toxicity can also affect treatment adherence. Newer agents are increasingly associated with lesser rates and severity of adverse drug effects; however, the user community remains highly sensitive to this central aspect of treatment that influences the quality of life. Information and beliefs can spread rapidly, and influential voices and vehicles of opinion may alter the perception of the community on treatment adequacy.

Filtering Tweets for the Study

The study that we present is based upon tweets filtered by specific keywords related to HIV and HIV treatments. Our work builds on a growing body of literature that uses combined human and computational approaches to assess health and disease dynamics from digital media [2,10]. Our goal was to determine if there are common adverse effects related to a particular HIV treatment, and to establish overall user sentiment (a positive, neutral, or negative perception) associated with the content of the tweet. Through our study, we defined methods to identify populations of interest. We used crowdsourcing (Amazon Mechanical Turk) to rate tweets to create training samples for our machine learning algorithms. On the analytical side, these algorithms were used to identify most of the tweets posted by our identified community (hereafter, referred to as “signal”): users with HIV whose tweets included references to treatments, symptomatic descriptions, opinions, and feelings. The final filtering of identified tweets was again obtained through crowdsourcing. Despite the limited number of tweets and individuals that may communicate publicly about HIV treatment experiences, we found a surprisingly large proportion of reports of drug toxicity, a high level of precision on drug-specific effects, and a general negative perception of treatment. We believe that monitoring of social media will be informative for the field, and broadly applicable to the surveillance of other therapies with long duration of use, even for those targeted at diseases that are associated with considerable stigma.

Methods

Dataset

For this study, we purchased 39,988,306 tweets posted between September 2010 and August 2013. We purchased the data from Gnip Inc, an official Twitter data reseller, which has recently been acquired by Twitter Inc. These tweets represent the full and unbiased stream of data during that period of time. A tweet was included in the dataset if at least one of the following keywords was contained within the tweet: *Sustiva*, *Stocrin*, *Viread*, *FTC*, *Ziagen*, *3TC*, *Epivir*, *Retrovir*, *Viramune*, *Edurant*, *Prezista*, *Reyataz*, *Norvir*, *Kaletra*, *Isentress*, *Tivicay*, *Atripla*, *Trizivir*, *Truvada*, *Combivir*, *Kivexa*, *Epzicom*, *Complera*, *Stribild*, *HIV treatment*, *HIV drug*, *anti-HIV*, *triple therapy HIV*, or *anti HIV*. These keywords were chosen in line with the listing of main antiretroviral agents in use [11]. *AZT* was dropped because of its extreme noise to signal ratio (as *azt* is one of the most used words in Hungarian). For technical reasons, the search query had to be limited in the numbers of search terms used.

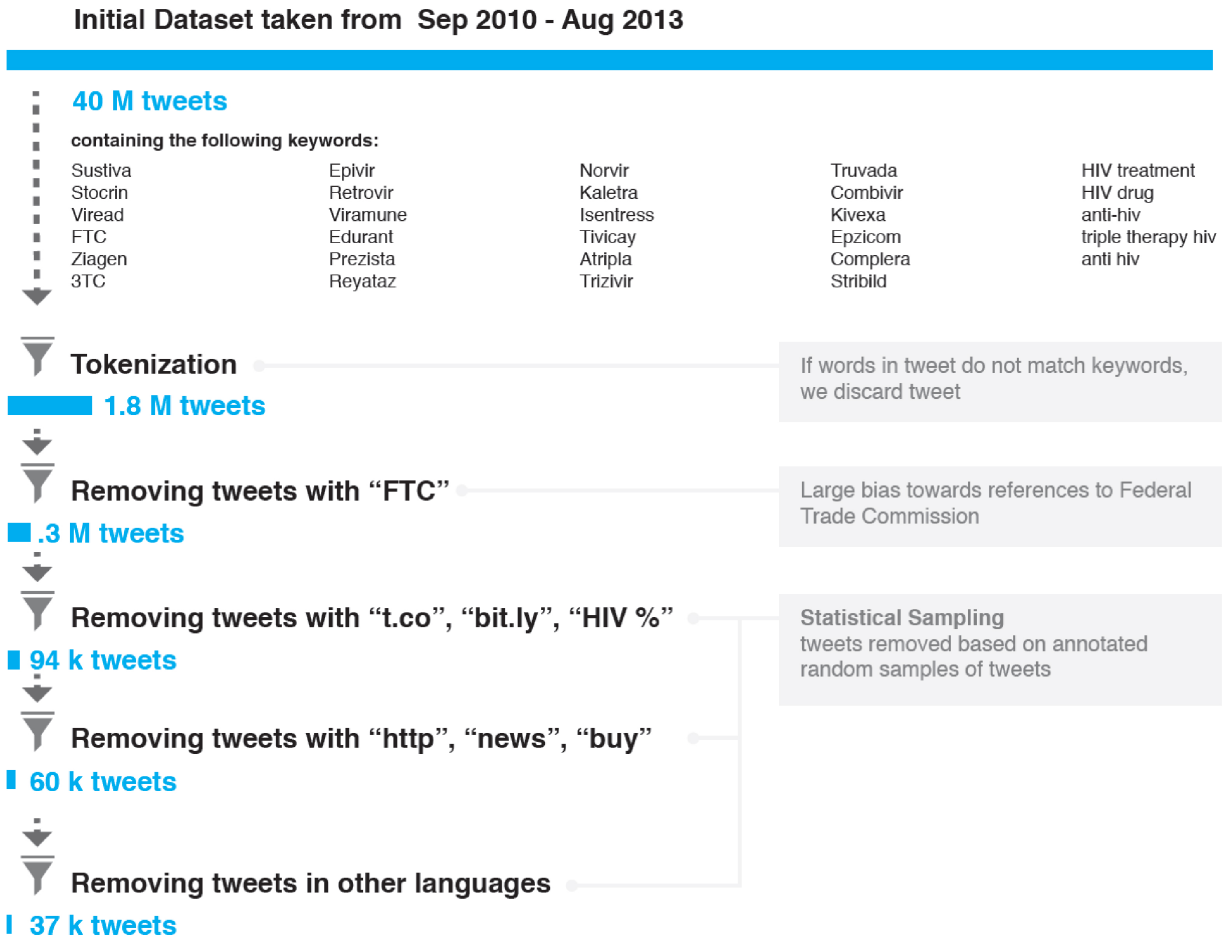
Processing of Tweets

For each collected tweet, we created a list of all the distinct tokens that the tweet contained. If at least one keyword matched at least one item in the list, we kept that tweet. This step reduced our data sample to about 1.8 million tweets, mainly due to the presence of compound words such as *giftcard*, triggered by the keyword *FTC*, which represented a large source of noise. Moreover, a subsample of tweets containing the keyword *FTC* presented a large bias toward information related to the Federal Trade Commission. We decided to discard this subsample from the data.

Hereafter, “signal” denotes our identified tweets posted by patients, and “noise” denotes tweets about unrelated topics, such as objective sentences concerning new discoveries or recent information about HIV in general.

Our statistical sampling analysis (Figure 1 shows this) started with loose criteria to reduce the noise in our sample. We used computational algorithms to randomly select different sets of tweets that contained words from potential noisy sources and sets of tweets that did not contain these “noisy” words. These sets of tweets were manually annotated by two of the authors (CA and TB) as either noise or signal. In the case of a slight indication of subjectivity, the tweet was annotated as signal. If the differences between the two types of sets (ie, containing or not containing specific noisy words) were significant in terms of signal content, we discarded tweets from our dataset that contained those noisy words (for more details, see below under the subheading “Identification of Signal Through Consecutive Filters”). Although we expected to discard some signal tweets through this cleaning process, it provided robustness to our procedure, as it allowed us to increase the percentage of signal tweets in the remaining sample. This higher purity of signal allowed a higher number of crowdsourced signal tweets to define training samples.

Figure 1. Overview of the different filters applied for the processing of tweets. M=million; k=thousand.



Identification of the Community

The goal of the data curation that we pursued was to purify the original noisy sample of tweets to a sample containing only signal. To do this, we defined a set of features that transformed a tweet into quantitative information (ie, number of words in the tweet, number of adverbs, etc). Moreover, these features were selected based on their separation power between objective information and subjective sentences charged with personal references. In a further step, we combined these features into a single output through a machine learning algorithm.

Training Samples

To define training samples for the classes of signal, noise, and non-English, we performed a crowdsourcing request of 4000 tweets. There were two Amazon Mechanical Turk workers that rated these tweets. We asked the two workers whether they believed the tweet fulfilled one of the following four criteria: (1) talking about personal medication; (2) talking about medication, but not personally; (3) talking about completely irrelevant topics to our study; and (4) tweet not in English. If both workers agreed on their respective answers concerning a given tweet, we kept the tweet for our analysis. Tweets rated as category (1) were used as signal, and categories (2) and (3) as noise. We had an agreement of about (3118/4000) 77.95% between workers that rated our tweets. Moreover, we used tweets rated as non-English as a control sample in order to remove foreign language tweets. We removed non-English tweets with a method described in see Multimedia Appendix 1.

Machine Learning Classifier

We utilized the Toolkit for Multivariate Analysis library [12] to define our machine learning classifier. We computed the signal efficiency versus noise rejection for four types of classifiers: (1) Boosted Decision Trees with AdaBoost (BDT), (2) Support Vector Machines (SVM), (3) Boosted Decision Trees with Bagging (BDTG), and (4) artificial neural networks. The figure of merit that we used to optimize our classifier was the noise rejection efficiency at a signal efficiency of 90%. We used this point on the receiver operating characteristic curve because it represents an optimal threshold to keep a large fraction of relevant signal tweets, while removing noise at a level that permits a final human validation. This further crowdsourcing was the final step in our identification of signal and was applied right after the machine learning algorithm.

Annotation of Side Effects and Sentiment Scores

An important aspect of analysis of Twitter data that cannot easily be obtained through other digital media (eg, search queries from Google or Bing) is the possibility to attribute an overall "sentiment". For this study, we hand-rated tweets on a sentiment scale ranging from -5 to 5, in steps of 1. The former indicates an extremely negative sentiment and the latter an extremely

positive sentiment. The following tweets are examples from the final dataset.

> *5: Hey guys I m officially undetectable!!!! Take that #hiv! CD4 went up bout 150 pts also! yaya!! #atripla*
>
> *4: Whoever invented Atripla (and it s component parts - it s three drugs in one) are geniuses. I love you.*
>
> *3: and more exciting and totally separate from this weekend, I am moving drug combinations from Atripla.*
>
> *2: I hear Truvada makes you fat like a Tellytubby. I am soooooo excited to finally be able to look like my hero, Tinky Winky!*
>
> *1: @TheBodyDotCom Taking Good food #nutrition as part of myother side effects.*
>
> *0: So,you better all get started taking Truvada (like I already do). It s OK, coz the FDA say so. hmmmm*
>
> *-1: #Atripla vs #HIV These meds seem to be getting short andwas still flying.*
>
> *-2: Atripla is a B**CH when you have to be up early in thefor 6 hours!*
>
> *-3: Oh boy here goes that feeling I hate.... #Atripla*
>
> *-4: I think I ll pass on the Atripla again today though. I feel weak but my future is so bleak I just want to waste away.*

(There were no tweets that we felt met out criteria of an extremely negative -5 rating.) Our measure of sentiment, Ψ, is the average of the sentiments of all tweets in a given time window, after having assigned a systematic uncertainty of 1 to each rating, which lead to a total uncertainty in the average of $1 / \sqrt{N}$. The sample of pure signal tweets was annotated by one of the authors (CA). Side effects were transcribed with no previous knowledge of drug toxicities.

Identification of Signal Through Consecutive Filters

After we applied the cleaning process described above, our dataset contained a larger fraction of possible signal tweets. In this section, we detail the steps taken to reach a sample that contained only signal tweets (Figure 1).

In a first step, we selected three random samples of 500 tweets containing tokens *t.co*, *bit.ly*, and starting with the term *HIV*. In Twitter, the tokens *t.co* and *bit.ly* are used as part of hyperlinks. We selected also one random sample of 500 tweets not containing these words. For the three filters, we found 0, 1, and 1 possible signal tweets in the first samples, respectively. There were 24 possible signal tweets that were found in the last sample where the tokens *t.co*, *bit.ly*, and starting with *HIV* were excluded. From this, we derived that 7.4 ± 2.7 possible signal tweets were expected to be found in 500 tweets taken from the original 316,081. The subsample with *t.co*, *bit.ly*, and starting with *HIV* contained a very low fraction of possible signal tweets, and was discarded from the dataset. Our findings indicate that these tokens appear within tweets addressing an objective idea.

In a second step, we checked tweets containing *http*, *news*, or *buy*. We selected three random samples of 140 tweets and found zero tweets in each of the three cases. As we found 24 possible signal tweets in a total of 500 tweets, we expected 6.7 ± 2.6 possible signal tweets in 140 tweets, assuming a Poisson distribution. We computed the probability to discard a possible signal tweet if we removed tweets containing *http*, *news*, or *buy* to be 4×10^{-6}. After we discarded tweets containing *http*, *news*, or *buy*, our dataset was reduced substantially.

In the final step, we separated two samples of 150 tweets containing *free*, *buy*, *de*, *e*, *za*, *que*, *en*, *lek*, *la*, *obat*, *da*, *majka*, *molim*, *hitno*, *mil*, or *africa*. These tokens were selected on the basis of forming part of foreign dictionaries, belonging to retailing companies selling their products, or seemingly forming part of news. We found zero possible signal tweets in both cases. Then, we estimated that the probability of losing possible signal tweets was less than 5×10^{-6}, if we removed tweets containing at least one of this large set of words. Following the aforementioned processing of tweets, the sample was further reduced to 37,337 tweets, or about 0.933% of the original sample size of 40 million tweets. Figure 1 presents a visualization of the various filters applied to our dataset in the cleaning process.

To identify tweets posted by our community of interest, we continued to filter the sample of 37,337 tweets. An SVM classifier, trained using the variables (see Multimedia Appendix 1) personalcount, tagnoun, sis noise, sis signal, bigrams noise, is english, common noise, common signal, and ncharacters, allowed to reduce the noise of the aforementioned sample by (481/603) 79.8%, while keeping (26/30) 87% of the signal tweets (see Multimedia Appendix 1, Figure S1). We were confident with the performances of our classifier, as the results obtained using a different testing (cross-check) sample agree within the statistical uncertainties. The classifier used was trained using 603 noise and 49 signal tweets, respectively. The performances of this classifier were tested and then validated using two different sample sets that contained each 603 noise and 30 signal tweets, respectively.

The output of our classifier is a real number between 0 and 1. Higher values indicate higher probability of being signal. In order to estimate the threshold to apply to our sample, we used the annotated signal tweets and computed the 90% signal efficiency threshold to be 0.45. Therefore, we parsed our entire sample of tweets through our classifier and kept only those tweets with classifier outputs larger than 0.45. As a result of this filtering, our remaining sample was reduced to 5443 tweets that we sent for crowdsourcing rating. Finally, after crowdsourcing, our pure sample of signal contained 1642 annotated tweets.

Results

Analysis of the Identified Community

The selection described above allowed the identification of users that tweet about their daily lives in the context of HIV. There were 512 unique users that posted the tweets identified as signal. We identified 247/512 (48.2%) male users, 83/512 (16.2%) female, and 182/512 (35.5%) that were unidentified gender. Gender was annotated manually either by first name (if conclusive) or by self-reported identification (if available). About half of users also identified their location. Half of these

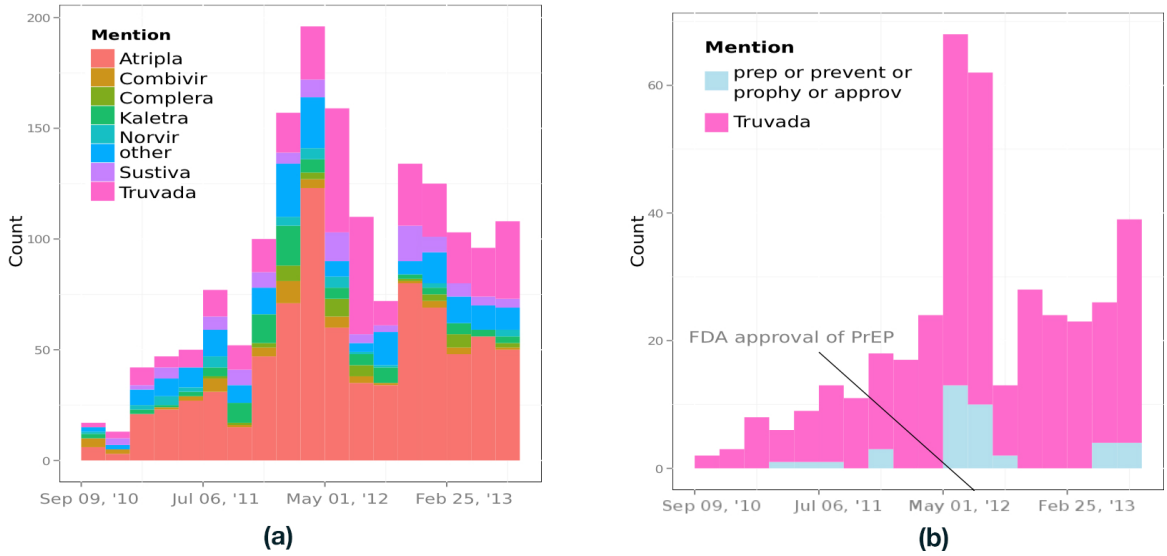
self-identified locations were in the United States, mainly from New York City and San Francisco. Most of the other locations were from countries with English as the official language, including the United Kingdom, South Africa, and Canada.

The identified community has a large average follower count of about 2300. When looking at the number of within-community followers, for example, the followers of each user who are also in the same community, we find that almost half of the users have no followers within the same community, and several users have almost 100 followers in the community. Furthermore, we studied the friendship relationship between our followers: Are the users that are followed by another member of the community following said follower?. The results indicate that indeed the distribution of these reciprocal relationships is very similar to the one only considering followers, which indicates the strong ties of friendship within our community. Nevertheless, the user friendship network is far from being fully connected, it represents 1516 edges and a structure of 245 subgraphs of connected components.

Trends in Antiretroviral Drug Tweets

Figure 2 shows (left image) a visualization of HIV drugs mentioned by users from September 9, 2010 through August 28, 2013. It shows a peak during the first 6 months of 2012. The total number of tweets is shared among all drugs equally in the first two bins. This trend disappears after the third bin, where Atripla receives more explicit mentions. The only period where Atripla is not ranked first in terms of mentions corresponds to the time spanning from May through August 2012, when Truvada, used as part of a strategy to reduce the risk of HIV infection, gained mentions. We evaluated in more detail the change of ranking of Truvada. Figure 2 displays (right image) the number of Truvada occurrences and the occurrences of four substrings combined: *prep*, *prevent*, *prophy*, and *approv*. The distribution of occurrences of these four tokens aligns with Truvada occurrences, indicating that we captured users tweeting about the approval of Truvada by the US Food and Drug Administration on July 2012 as prophylaxis.

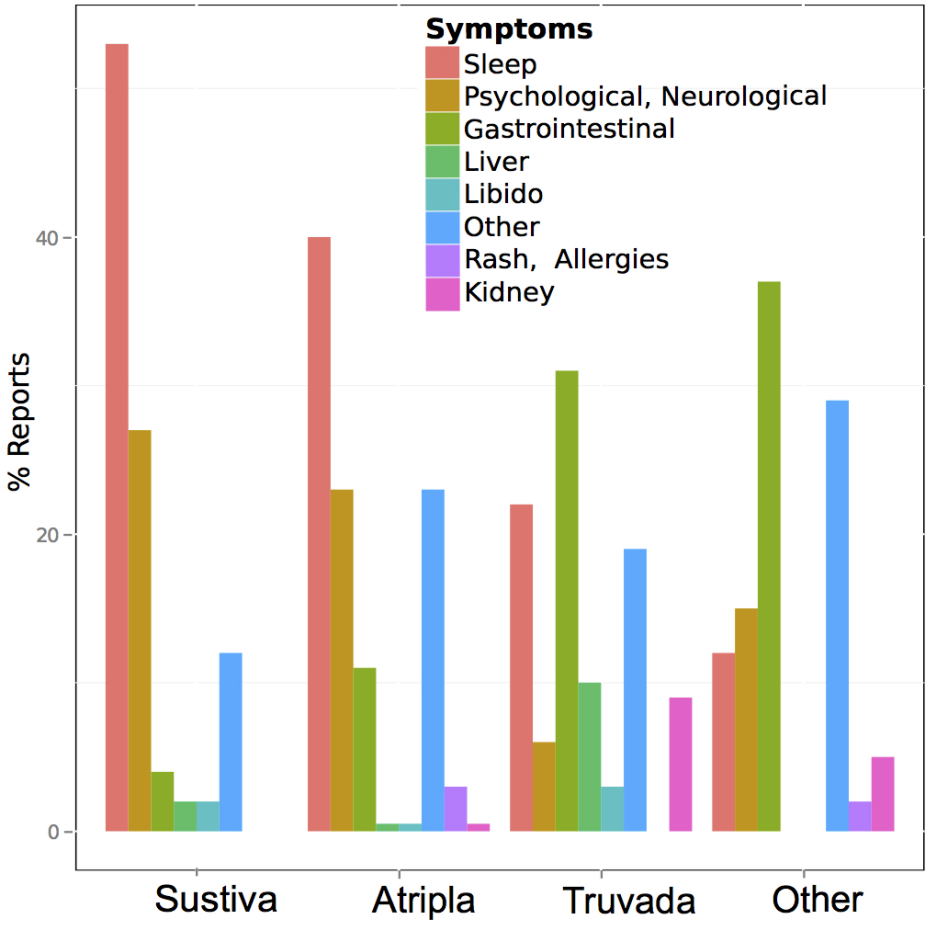
Figure 2. Number of tweets that contain specific mentions as a function of time, from September 2010 until August 2013. Each bin spans a total of 60 days. Panel (a) notes the seven most tweeted drugs separately, and groups the rest of drug mentions under the label "Other". Panel (b) highlights the increase in Truvada tweets at the time of Federal Drug Administration approval of Pre-exposure prophylaxis (PrEP). Specific tokens (on blue) support the association.



Reporting on Drug Adverse Effects

Not considering retweets (ie, reposted tweets by other users), 329 out of 353 tweets contained precise information that could be captured as drug adverse effects. Corresponding to the high frequency of use of efavirenz and efavirenz-containing combination treatments (ie, Sustiva, Atripla), most users report problems regarding their sleep, be it a nightmare or a vivid dream, or lack of sleep, as well as symptoms comparable to the effect of psychoactive drugs. These are known adverse effects that may lead to treatment discontinuation [13]. The adverse effects of tenofovir, a component of the commonly used fixed-dose combination pill (Atripla), and of the post exposure prophylaxis regimens (Truvada), is associated with renal toxicity [14] and, in particular in the post exposure setting, with nausea and vomiting [15]. Other drug regimens, such as those including protease inhibitors, are commonly associated with gastrointestinal intolerance. Also, about (27/353) 7.6% of the selected relevant tweets indicate no side effects and that a specific drug was well tolerated. The summary representation of adverse effects attributed to specific drugs or drug combination accurately captures well-recognized toxicities [8,16] (Figure 3 shows this).

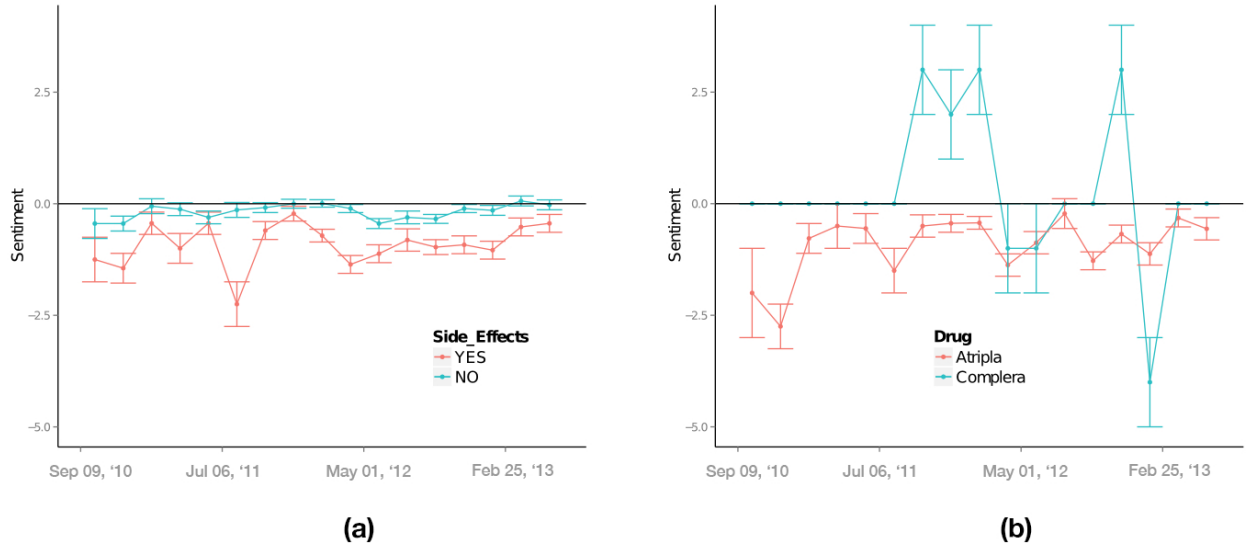
Figure 3. Summary of reported toxicities by unique users on Twitter between September 9, 2010 and August 23, 2013.



Analysis of User Sentiment

Central to this work was the analysis of the observed dynamics of tweets under the concept of sentiment, for example, the expressed emotion associated with the tweet content. Figure 4 shows the computed Ψ over the studied time. We captured an average negative sentiment of -0.178; from a total of 1347 tweets, 348 were rated as negative sentiment, 220 as positive, and 772 as neutral. We also assessed the sentiment associated with tweets specifically referring to adverse effects (Figure 4, left image). From 357 tweets referring to side effects, 238 were associated with negative sentiments, 78 positive, and 37 as neutral. Around May 1, 2012, the number of tweets associating adverse effects peaked at 54, while the average number was 20. This fact was paired with a negative sentiment at that time. Tweet sentiment dynamics appear to differentiate across specific drugs, as illustrated in Figure 4 (right image) with the comparison between Atripla (259 sentiment ratings) and Complera (8 sentiment ratings).

Figure 4. Sentiment score Ψ as a function of time. Panel (a) shows sentiment distributions obtained by considering all tweets and tweets with references to side effects, respectively. Panel (b) depicts sentiment distributions for tweets referring to Atripla and Complera, respectively. The uncertainties are estimated as referred in the Methods section.



Discussion

Using Social Media Data for Public Health

There has been increasing interest on the use of social media data for public health and medicine. Digital epidemiology [1,17] utilizes data from sources such as search engines (Google, Bing, etc), public Internet resources such as Wikipedia [18], and social online networks to track the dynamics of health and disease. Specifically, data from Twitter has been used for the analysis of influenza epidemics [19,20] and influenza vaccine sentiments. More recently, there has been increasing attention to the potential of Internet-based analytics for post marketing assessment of drug toxicity [2,6]. Analysis data from search engines can reconstruct known toxicities based on users queries. Importantly, pharmacovigilance using search query data can expand the spectrum of adverse effects for a given drug or drug interaction on the basis of "guilt by association" [21].

There is little scientific experience in the analysis and use of social media data for HIV. This contrasts with the clear interest that this approach could have for the understanding of global or local epidemiology, for social behavioral studies, and for research on prevention. A study by Young et al [22] served to explore social media for identification of HIV risk through tweets. From more than 550 million tweets, they extracted 9800 geolocated tweets that mapped to areas of high HIV prevalence. On the basis of their study, it has been suggested that an early warning indicator for HIV incidence could emerge by combining HIV prevalence and Twitter data [23].

We aimed here to conduct a first study, to our knowledge, on the possibility that HIV-infected individuals would tweet about their disease, and more specifically, about their experience with drug treatment. It is probable that the specific analysis of drug toxicity, and in particular the possible identification of expanded spectrum of adverse effects, could have been more efficient and successful through the study of queries in search engines such as Google or Bing. However, exploring these questions via analysis of Twitter data offered two unique features: (1) the possibility to capture the emotional context of the tweet, and (2) the estimate of the number of individuals (followers on Twitter) that could be influenced by the content of those tweets.

Extracting Information From Tweets

Extracting information from tweets requires a significant investment in filtering strategies to separate signal from noise. We placed particular attention on removing tweets that were not written by the identified population (ie, the community of infected individuals). We also went through a number of validation steps that included crowdsourcing and human intervention to assure the validity of filters and support identification of toxicities and hand-rating of sentiments. After applying data cleaning, machine learning, and human rating to an initial dataset of 40 million tweets, we identified 1642 tweets and 512 unique users disclosing personal information in the context of HIV. Although these numbers are not large, individuals that do tweet about their HIV drugs are heavily followed on Twitter, with an average number of followers of almost 2300. Analysis of the target community also revealed strong ties of friendship. This sizable landscape of contacts and possible influence becomes particularly relevant when we examined the sentiment associated with their treatment, as those perceptions can spread broadly to influence the community. As expected, relevant hubs were cities like New York and San Francisco.

The interest of Twitter and of social online networks rests on the fact that information has a personal character that can be captured as "sentiment". Tweets can be tagged as expressing a positive, negative, or neutral sentiment. About (137/531) 25.8% of tweets on drugs were scored as "negative". Up to (424/1491) 28.43% of the tweets concerned adverse drug effects, and (281/424) 66.2% of those were scored as negative. It was clear from the data that the concerns expressed in those tweets were accurate as to the nature of the toxicities associated with different components of the anti-HIV combination treatment.

For example, the description of adverse effects of efavirenz-containing drugs (Sustiva, Atripla) reconstructs well-established patterns of the neuropsychological toxicity of this compound [8]. The analysis also allowed capturing the rates of reporting, and the associated sentiment in time among those that described side effects. These analyses discriminated across drugs in terms of general acceptance (negative vs neutral-positive sentiments) over time.

The analysis also identified the response of the community to unique events. In Spring 2012, we observed a peak of tweeting about Truvada, at the time that this drug was approved by the Federal Drug Administration for use as post exposure prophylaxis. Although increased media attention would result in increased Twitter communication, we observed that increasingly negative sentiment scores accompanied the increase in tweets, including for tweets referring to adverse effects. This underscores the interest for public health stakeholders promoting new preventive measurements to track these potentially influential signals from online social networks.

The limits of the study are derived from the low numbers of individuals actually reporting on HIV drugs through Twitter. However, as discussed above, the reporting of toxicities is accurate in regards of what is known about those drugs. Different also from search engines, where individuals are actively searching for information possibly related to new adverse events, the content of tweets may simply reflect the awareness of symptoms and signs that have already been announced by health care professionals or in Web resources. The reporting may also reflect preconceived notions of shared beliefs from the community. This observation is in line with the work of Freifeld et al [6] that analyzed 4401 tweets of potential adverse events to 23 medical products and found a r=0.75 correlation with consumer-reported Federal Drug Administration Adverse Event Reporting System reports at the System Organ Class level. However, whatever is reported, it indeed reflects on contemporary sentiments among users and their followers. Moreover, previous studies using Twitter were limited by only accessing a small fraction of the tweets. In contrast, we acquired from a specialist service all tweets containing the keywords for the study interval, and we therefore have a full representation of the data for the period of the study. There is also increasing skepticism regarding the use of online social media data as the sole source of information. There has been considerable debate on the recent failure of Google Flu to accurately match influenza trends as compared with the sentinel data from the US Center for Disease Control and Prevention [24]. We believe that our approach that combines automated filters, crowdsourcing, machine learning, and extensive manual validation and scoring may bring more reliable results, even if at the cost of lesser automation. The approach would apply and be a useful addition to real time surveillance of other therapeutic interventions at the population level.

Acknowledgments

MS gratefully acknowledges the National Institutes of Health Research and Policy for Infectious Disease Dynamics support. AT received support from the Santos Suarez Foundation for this work.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Method used to remove non-English tweets.

[PDF File (Adobe PDF File), 176KB - publichealth_v1i2e7_app1.pdf]

References

1. 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]
2. Salathé M, Khandelwal S. Assessing vaccination sentiments with online social media: Implications for infectious disease dynamics and control. PLoS Comput Biol 2011 Oct;7(10):e1002199 [FREE Full text] [doi: 10.1371/journal.pcbi.1002199] [Medline: 22022249]
3. 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]
4. Pagoto S, Schneider KL, Evans M, Waring ME, Appelhans B, Busch AM, et al. Tweeting it off: Characteristics of adults who tweet about a weight loss attempt. J Am Med Inform Assoc 2014;21(6):1032-1037. [doi: 10.1136/amiajnl-2014-002652] [Medline: 24928175]
5. Jashinsky J, Burton SH, Hanson CL, West J, Giraud-Carrier C, Barnes MD, et al. Tracking suicide risk factors through Twitter in the US. Crisis 2014;35(1):51-59. [doi: 10.1027/0227-5910/a000234] [Medline: 24121153]
6. Freifeld C, Brownstein J, Menone C, Bao W, Filice R, Kass-Hout T, et al. Digital drug safety surveillance: Monitoring pharmaceutical products in twitter. Drug Saf 2014 May;37(5):343-350 [FREE Full text] [doi: 10.1007/s40264-014-0155-x] [Medline: 24777653]

7. Sarker A, Ginn R, Nikfarjam A, O'Connor K, Smith K, Jayaraman S, et al. Utilizing social media data for pharmacovigilance: A review. J Biomed Inform 2015 Apr;54:202-212 [FREE Full text] [doi: 10.1016/j.jbi.2015.02.004] [Medline: 25720841]
8. Fellay J, Boubaker K, Ledergerber B, Bernasconi E, Furrer H, Battegay M, Swiss HIV Cohort Study. Prevalence of adverse events associated with potent antiretroviral treatment: Swiss HIV Cohort Study. Lancet 2001 Oct 20;358(9290):1322-1327. [Medline: 11684213]
9. Keiser O, Fellay J, Opravil M, Hirsch H, Hirschel B, Bernasconi E, Swiss HIV Cohort Study. Adverse events to antiretrovirals in the Swiss HIV Cohort Study: Effect on mortality and treatment modification. Antivir Ther 2007;12(8):1157-1164. [Medline: 18240856]
10. Sadilek A, Brennan S, Kautz H, Silenzio V. nEmesis: Which restaurants should you avoid today?. Palm Springs, California; 2013. URL: http://www.cs.rochester.edu/u/kautz/papers/Sadilek-Brennan-Kautz-Silenzio_nEmesis_HCOMP-13.pdf [accessed 2015-07-14] [WebCite Cache ID 6a1tfZoUa]
11. Günthard HF, Aberg JA, Eron JJ, Hoy JF, Telenti A, Benson CA, International Antiviral Society-USA Panel. Antiretroviral treatment of adult HIV infection: 2014 recommendations of the International Antiviral Society-USA Panel. JAMA 2014;312(4):410-425. [doi: 10.1001/jama.2014.8722] [Medline: 25038359]
12. Hoecker A, Speckmayer P, Stelzer J, Therhaag J, Toerne VE. TMVA - Toolkit for Multivariate Data Analysis. 2007. TMVA - Toolkit for multivariate data analysis URL: http://arxiv.org/abs/physics/0703039 [accessed 2015-07-01] [WebCite Cache ID 6ZhC3T3yO]
13. Ford N, Shubber Z, Pozniak A, Vitoria M, Doherty M, Kirby C, et al. Comparative safety and neuropsychiatric adverse events associated with efavirenz use in first-line antiretroviral therapy: A systematic review and meta-analysis of randomized trials. J Acquir Immune Defic Syndr 2015 Aug 1;69(4):422-429. [doi: 10.1097/QAI.0000000000000606] [Medline: 25850607]
14. Ryom L, Mocroft A, Lundgren JD. Antiretroviral therapy, immune suppression and renal impairment in HIV-positive persons. Curr Opin HIV AIDS 2014 Jan;9(1):41-47. [doi: 10.1097/COH.0000000000000023] [Medline: 24225381]
15. Coutinho B, Prasad R. Emtricitabine/tenofovir (Truvada) for HIV prophylaxis. Am Fam Physician 2013 Oct 15;88(8):535-540 [FREE Full text] [Medline: 24364575]
16. Fagard C, Le BM, Günthard H, Hirsch HH, Egger M, Vernazza P, Swiss HIV Cohort Study. A controlled trial of granulocyte macrophage-colony stimulating factor during interruption of HAART. AIDS 2003 Jul 4;17(10):1487-1492. [doi: 10.1097/01.aids.0000060396.18106.5b] [Medline: 12824786]
17. Salathé M, Freifeld CC, Mekaru SR, Tomasulo AF, Brownstein JS. Influenza A (H7N9) and the importance of digital epidemiology. N Engl J Med 2013 Aug 1;369(5):401-404. [doi: 10.1056/NEJMp1307752] [Medline: 23822655]
18. McIver DJ, Brownstein JS. Wikipedia usage estimates prevalence of influenza-like illness in the United States in near real-time. PLoS Comput Biol 2014 Apr;10(4):e1003581 [FREE Full text] [doi: 10.1371/journal.pcbi.1003581] [Medline: 24743682]
19. 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]
20. Kim E, Seok JH, Oh JS, Lee HW, Kim KH. Use of hangeul twitter to track and predict human influenza infection. PLoS One 2013;8(7):e69305 [FREE Full text] [doi: 10.1371/journal.pone.0069305] [Medline: 23894447]
21. White R, Tatonetti N, Shah N, Altman R, Horvitz E. Web-scale pharmacovigilance: Listening to signals from the crowd. J Am Med Inform Assoc 2013 May 1;20(3):404-408 [FREE Full text] [doi: 10.1136/amiajnl-2012-001482] [Medline: 23467469]
22. Young SD, Rivers C, Lewis B. Methods of using real-time social media technologies for detection and remote monitoring of HIV outcomes. Prev Med 2014 Jun;63:112-115 [FREE Full text] [doi: 10.1016/j.ypmed.2014.01.024] [Medline: 24513169]
23. Stoové MA, Pedrana AE. Making the most of a brave new world: Opportunities and considerations for using Twitter as a public health monitoring tool. Prev Med 2014 Jun;63:109-111. [doi: 10.1016/j.ypmed.2014.03.008] [Medline: 24632229]
24. 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]

Abbreviations

BDT: Boosted Decision Trees with AdaBoost
HIV: human immunodeficiency virus
SVM: Support Vector Machines

Edited by G Eysenbach; submitted 31.03.15; peer-reviewed by C Freifeld, N Zhang; comments to author 17.04.15; revised version received 29.04.15; accepted 30.05.15; published 27.07.15.

Please cite as:
Adrover C, Bodnar T, Huang Z, Telenti A, Salathé M
Identifying Adverse Effects of HIV Drug Treatment and Associated Sentiments Using Twitter
JMIR Public Health Surveill 2015;1(2):e7
URL: http://publichealth.jmir.org/2015/2/e7/
doi:10.2196/publichealth.4488
PMID:27227141

Original Paper

Electronic Cigarette Marketing Online: a Multi-Site, Multi-Product Comparison

Kar-Hai Chu[1], PhD; Anupreet K Sidhu[1], MA; Thomas W Valente[1], PhD

University of Southern California, Department of Preventive Medicine, Los Angeles, CA, United States

Corresponding Author:
Kar-Hai Chu, PhD
University of Southern California
Department of Preventive Medicine
2001 North Soto Street, 3rd Floor
Los Angeles, CA, 90032
United States
Phone: 1 323 442 7244
Fax: 1 323 442 8201
Email: karhaich@usc.edu

Abstract

Background: Electronic cigarette awareness and use has been increasing rapidly. E-cigarette brands have utilized social networking sites to promote their products, as the growth of the e-cigarette industry has paralleled that of Web 2.0. These online platforms are cost-effective and have unique technological features and user demographics that can be attractive for selective marketing. The popularity of multiple sites also poses a risk of exposure to social networks where e-cigarette brands might not have a presence.

Objective: To examine the marketing strategies of leading e-cigarette brands on multiple social networking sites, and to identify how affordances of the digital media are used to their advantage. Secondary analyses include determining if any brands are benefitting from site demographics, and exploring cross-site diffusion of marketing content through multi-site users.

Methods: We collected data from two e-cigarette brands from four social networking sites over approximately 2.5 years. Content analysis is used to search for themes, population targeting, marketing strategies, and cross-site spread of messages.

Results: Twitter appeared to be the most frequently used social networking site for interacting directly with product users. Facebook supported informational broadcasts, such as announcements regarding political legislation. E-cigarette brands also differed in their approaches to their users, from informal conversations to direct product marketing.

Conclusions: E-cigarette makers use different strategies to market their product and engage their users. There was no evidence of direct targeting of vulnerable populations, but the affordances of the different sites are exploited to best broadcast context-specific messages. We developed a viable method to study cross-site diffusion, although additional refinement is needed to account for how different types of digital media are used.

(JMIR Public Health Surveill 2015;1(2):e11) doi:10.2196/publichealth.4777

KEYWORDS

electronic cigarettes; content analysis; social networking sites; marketing

Introduction

Electronic cigarettes, also known as e-cigs or e-cigarettes, are battery-operated products that deliver nicotine by turning it into an aerosol that is inhaled by the user [1]. Today, awareness of e-cigarettes is high [2], and their use continues to grow [3,4]. With declining cigarette sales and the potential for increasing Food and Drug Administration (FDA) regulation, there is a rise in noncombustible tobacco products, with e-cigarette advertising being the most extensively circulated [5]. E-cigarette brands have been using widespread advertising campaigns, spending $541.7 million in 2011 [6].

The Internet represents a medium that offers an opportunity for e-cigarette brands to expand their audience reach. However, this unregulated domain can also be a cause for concern. Several studies have suggested that e-cigarettes can be a viable method to quit smoking [7,8], although Grana and Ling [9] found that health claims and smoking-cessation messages on many

websites, aided by images of doctors or celebrities, are unsupported by scientific evidence. Misinformation can be easily spread in many arenas online, and social networking provides a convenient and cost-effective venue for e-cigarette promotions of non-scientifically supported claims.

There are countless social networking sites with unique technological features and user demographics that can be attractive for selective marketing. The growth of e-cigarettes occurred as social media and other Web 2.0 sites became an important platform for commercial advertising. It is no surprise that e-cigarette brands, especially smaller ones that have no affiliation with larger tobacco companies and no sizable advertising budgets, took to social media to market their products. Recent e-cigarette studies have typically examined product health effects (eg, [10,11]). Few studies, however, have examined e-cigarette marketing within social media. In one example, Huang et al [12] found that e-cigarette tweets were mostly commercial (90%), and were posted by a small group of active accounts. They concluded that Twitter served as an important platform for e-cigarette marketing.

Social networking sites such as Facebook or Twitter are inherently built to support social networks. Whether posting messages to a Facebook wall or sending updates on Twitter, interactions between users are the foundation of all activity on these sites. The distinction in how users of different sites interact depends on the digital media affordances the sites focus on. Affordances are the actionable relationships between an actor and an aspect of the environment that offer the actor the potential for action [13]. This concept has been extended to human-computer interactions and applied to product designs, such as graphical user interfaces [14]. The concept of affordances is especially important as new technologies continue to be developed.

Web 2.0 tools provide an abundance of ways to interact, and social networking sites use these tools to great advantage. Kietzmann and colleagues' well-cited work [15] describes seven functional blocks of social networking sites: identity, conversations, sharing (of content), presence, relationships, reputation, and groups. Each site varies in the degree to which they use or promote these functions. For example, every Instagram post has a picture, and sharing is at the core of the site. Twitter, on the other hand, allows users to micro-blog by producing 140 character tweets that promote sharing and conversations. Facebook and Google+ combine various digital media, wrapped together to create a general social networking platform, focusing on relationships and identity. Indeed, Facebook friends are not the same as Instagram friends, and spreading messages on Twitter is different than doing the same on Google+. These differences in digital media afford unique ways by which users on these sites interact, and can result in different kinds of relationships being developed [16].

Social networking sites also have varying user demographics. A Pew 2013 report on social media [17] revealed that Facebook has an increasingly older population (users 65+), Twitter has high adoption by African Americans, women are 4 times more likely than men to be Pinterest users, Instagram is also popular among women, and LinkedIn is popular for higher-income households. Research using different methods of collecting demographic data, such as connecting self-reported user names with ethnicity based on US Census data, report similar results for Twitter [18] and Facebook [19]. There are other demographic trends, but it is clear that each social networking site has different levels of usage and appeal for certain demographics. These variations in social networking site demographics allow marketing companies to customize their advertisements and possibly target sub-populations based on their social networking tendencies.

The unique affordances and demographics of each social networking site, especially how they are used by online marketers, is the focus of this paper. In this study, we explore the possibility of e-cigarette brands potentially taking advantage of social networking site demographics and targeting vulnerable sub-populations, such as underage youth. We will also see if and how e-cigarette brands utilize different digital media in their marketing strategies. As such, we ask two primary research questions: (1) Are e-cigarette brands exploiting the affordances of each site in their marketing? And (2) Are e-cigarette brands targeting sub-populations (eg, women, teenagers) by taking advantage of the demographic differences of different social networking sites?

We also study how different social networking sites potentially interact with one another, for example when posting a tweet to one's Facebook wall. To our knowledge, no studies have examined the diffusion of information across multiple social networking sites. However, Pew's latest (2014) survey indicates that 52% of online adults use two or more social media sites, rising from 42% in 2013 [20]. Given our plans to collect information across multiple social networking sites, we are also interested in determining if viable methods can be developed to identify how users and information might cross the boundaries between different platforms. There are many implications in this line of study, including diffusion of messages, how different technologies can change message content, and roles that people might serve in bridging social networks from different sites, to name a few. This leads to our final exploratory research question: Are there identifiable instances of users and information crossing different social networking sites?

Methods

Data

This study examined four social networking sites and two e-cigarette brands that maintain a presence in each site for marketing and advertising. Data from each of the four sites were collected from October 20, 2012 through April 14, 2015. The sites were chosen based on popularity (from current eBizMBA.com rankings), differences in the affordances of the platform's technologies, accessibility of data, and user demographics [20,21]. The final set of sites included Facebook (general purpose, with increase usage among seniors), Twitter (text micro-blog, with high adoption by African Americans), Google+ (general purpose), and Instagram (picture based, with high adoption by women). E-cigarette brands were chosen based on activity in multiple social networking sites and general online presence. The two choices were Blu (owned by Lorillard, then

by Reynolds Tobacco, and finally by Imperial Tobacco in 2014) and V2 (owned by VMR). Currently, Blu and V2 are two of the top three e-cigarette brand websites based on activity, according to traffic tracker Compete.com, with similarly high social media activity. Blu also has the highest advertising expenditures, comprising more than 75% of all e-cigarette advertising expenditures in 2012 [22]. These two choices also allow us to see potential differences between a tobacco industry-owned brand (Blu) and a privately owned brand (V2).

Data were collected by establishing connections to each social networking site via the platform's Application Programming Interface (API). An API allows external software to make requests for data and post information and to perform tasks that are made allowable by the service. Custom software was written to communicate with the API of each of the four social networking sites. The 3 main objectives were to:

1. Gather any available long-term historical data, primarily any content posted by Blu or V2 and related information, such as users who shared or commented on such posts. This process was conducted a single time.
2. Gather recent data as it was being posted. This task was the same as objective 1 but only collected recent data (about 3-4 days old), as many social networking sites limit the amount of data that can be requested. This process was executed every 3-4 days.
3. Gather related data unique to each site, as their methods of online interactions are different. For example, the data collection included retweets on Twitter, comments on Facebook, shares on Google+, likes on Instagram, and so on. Users who performed these actions were also recorded, along with any available demographic information. This process was executed every 3-4 days, in parallel to objective 2.

All data collected were publicly available since any person with an Internet connection is able to view data that has been retrieved through this mining software. The primary data being collected were content posted on the social networking sites. Demographic information came from users' self-reported data in the site and was only collected if the user had made it public. Personal information such as emails, phone numbers, or addresses was not collected. This study was reviewed by the University of Southern California Health Sciences Institutional Review Board, which found that it does not qualify as Human Subjects Research and thus is not subject to the requirements of 45 CFR 46.102.

Analysis

We analyzed the content posted on each of the social networking sites and parsed data according to e-cigarette brand and social networking site. Our primary method of analysis was Term Frequency-Inverse Document Frequency (TF-IDF) weighting a commonly used statistical method in information retrieval to calculate word relevance for a document across a large corpus [23]. TF-IDF classifies documents by examining each word and calculating its occurrence frequency. TF is the measure of importance of a single term in one document, as determined by its frequency (usually normalized by the total number of words in the document). The TF score is balanced by the IDF, which counts the number of documents a term appears in across the entire corpus, and takes the inverse. Thus, IDF measures the discriminating power of a term. In short, TF-IDF classifies any single document by terms that frequently appear in it but do not appear in many other documents. More advanced techniques can also build from this foundation, although they were not used in this study. We used the Simstat 2.6.2 software package and its associated content analysis component, WordStat 6.1.23, to support the analyses.

We conducted an exploratory network analysis of users who had posted content or interacted with one of the e-cigarette brands in one of the four social networking sites. To examine the possibility of cross-platform diffusion of e-cigarette messages, we ran the test in two steps. First, we identified usernames from each site that had an identical matching username in another site. While this tactic did not guarantee certainty that the different users are the same, it provided a starting approximation. We then analyzed content posted by the identified users in the different sites. Using the names that were found in multiple sites, we confirmed it was in fact the same person via identical profile pictures or personal descriptions, and then examined whether their activity on one site (eg, liking a Blu post on Facebook) resulted in sharing the content on another site (eg, posting the same message on Twitter).

Results

Table 1 describes the data, including total counts of posted content on each of the four social networking sites. Table 2 is a list of the top 20 terms used based on TF-IDF score. These are the top terms based on frequency and number of cases they appear in (representing how unique they are for each possible brand/site combination). The terms help contextualize the cases. As we are studying four social networking sites and two e-cigarette brands, there are 8 possible brand/site scenarios, or cases, that any given term can appear in. For example, the term *ecig* is used by both companies in all four sites, and thus has a count of 8 for number of cases, also called an 8-case scenario. In another example, the term *RT* is used by both companies but only on Twitter, and thus has a count of 2 for number of cases.

Table 1. Summary of posted data collected.

	Facebook	%	Google+	%	Instagram	%	Twitter	%
Total	9915		371		638		7861	
Blu	6313 (2106) [a]	64% (61%)	335	90%	342	54%	4753	60%
v2	3602 (1343) [a]	36% (39%)	36	10%	296	46%	3108	40%

[a]Facebook data includes posts by other users to be displayed on the e-cigarette brand page. Number in parenthesis represents posts made by the e-cigarette brands, ie, Blu and V2.

Table 2. Top terms based on TF-IDF scores.

	Frequency	% Total	No. cases	% cases	TF-IDF
RT	1200	0.50%	2	25.00%	722.5
BLUCIGS	1827	0.80%	4	50.00%	550
CO	4111	1.80%	6	75.00%	513.6
HREF	559	0.20%	2	25.00%	336.6
BLU	1302	0.60%	5	62.50%	265.8
NOFOLLOW	409	0.20%	2	25.00%	246.2
REL	409	0.20%	2	25.00%	246.2
BLUCRM	272	0.10%	1	12.50%	245.6
DM	325	0.10%	2	25.00%	195.7
OT	410	0.20%	3	37.50%	174.6
BLUNATION	556	0.20%	4	50.00%	167.4
CLASS	466	0.20%	4	50.00%	140.3
SAVE	503	0.20%	5	62.50%	102.7
PWD	110	0.00%	1	12.50%	99.3
WARD	163	0.10%	2	25.00%	98.1
VAPORIZER	206	0.10%	3	37.50%	87.7
SXSW	202	0.10%	3	37.50%	86
HASHTAG	412	0.20%	5	62.50%	84.1
BLUFREEDOM	275	0.10%	4	50.00%	82.8
HTTPS	589	0.30%	6	75.00%	73.6

In contrast, a 1-case scenario shows words that are only found in a single brand/site case (eg, *BLUCRM* or *PWD* in Table 2). In other words, those terms were only used by a single company on a single site. These terms are helpful in explaining why only one brand might be applying a specific marketing strategy on a single site. Applied to only a single case out of eight possibilities, it is the most conservative classification of the case by the terms found. These terms are unique, based on either the brand's usage or how the technology supports certain features. We found 163 terms that fit the 1-case scenario and classified each according to 15 possible categories. The categories were developed through an exploratory examination of the data, and the terms were coded by two of the authors. The authors agreed on the categories of 121 of the terms (74%). The remaining disagreements in classifications were discussed until a mutual agreement had been reached for all 163 terms. The most frequent categories found in the data were: 68% of Blu's terms on Twitter were for user interactions; 84% of Blu's terms on Facebook were for political information; and 73% of V2's terms on Twitter were links to their homepage. No strong content data were found to help classify either brand's activities on Google+ or Instagram, or for V2 on Facebook.

The results showed that Blu and V2 had contrasting strategies in their social networking site presence. In the 1-case data, V2 focused primarily on Twitter (94% of all of V2's 1-case terms), with the majority of interactions aimed at connecting followers to their home website. V2 did not have any notable 1-case discriminating content on Facebook, Google+, or Instagram. On Twitter, V2's content remained focused on website advertisement, although interactions with users were also included. Blu had unique content in both Twitter (66%) and Facebook (28%). Unlike V2, Blu's use of Twitter focused on interacting with users on a wide range of topics, from product support to general conversation. On Facebook, Blu's posts centered on political activities (eg, suggestions to email state representatives or city council members, information on rallies,

etc). Blu had no significant unique activity on Google+ or Instagram. Here are content examples from several of the top categories:

(Blu/Twitter/User interaction) @anonymized_username but that doesn't say, "I'm either really embarrassed or really proud to wear this at Christmas".

(Blu/Twitter/Event info) We've seriously had Fun Fun Fun giving out these @funfunfunfest tix #FFFfest

(Blu/Facebook/Political info) Did you see where a local Wisconsin legislature wants to pass a bill ensuring e-cigs are allowed to be used in public? Hit LIKE if you support this!

(V2/Twitter/Website) If you want to try some NEW flavors on your V2 battery, check out our clearance section!

(V2/Twitter/User interaction) @anonymized_username Nice V2 stash! Happy vaping! :)

A second follow up was conducted to investigate the terms in each brand/site with the highest raw frequency by removing any discriminating factor based on IDF. In addition, we filtered out site-specific terms (eg, *RT* in Twitter or *href* in Google+) to focus on terms that are topic-specific (ie, focused on e-cigarettes). Without the IDF discriminator or site-specific terms, these data provide a broad view of all general-purpose terms that are used in each site, according to each brand. We were able to see what themes and concepts the e-cigarette brands were broadcasting to their followers, regardless of which site they were using. The results are shown in Table 3.

Table 4 is a matrix showing the number of users that share a presence between two given sites. The diagonal shows the total number of users that had activity within a single site. A small percentage of matched screen names were found between profiles of users in the different sites. In terms of raw frequency, Instagram/Twitter had the most matches, followed by Instagram/Facebook, and Twitter/Facebook. We selected the top three users found in each site case, with six possible cases (three pairs of two), for 18 potential users. All users were confirmed to be the same across sites by either identical images or profile descriptions. In each of the 18 cases, we found that no identical content crossed site boundaries.

Table 5 shows non-content interactions between users and an e-cigarette brand, separated into three types: 1) *comments*, which include text responses to posted content, 2) *likes/plusoners*, which includes a single supporting action that is collectively aggregated, and 3) *resharers/retweets*, which are actions where users repost existing content. These actions were not fully inclusive, as we only collected those that were relevant to this study.

Table 3. Top terms by raw count, with website coding terms removed.

Blu				V2			
Facebook	Google+	Instagram	Twitter	Facebook	Google+	Instagram	Twitter
Blu	vaping	blucigs	blucigs	cigs	products	cigs	cigs
blucigs	blunation	blunation	blu	shop	cig	vapor	new
cigarettes	blucigs	vaping	blucrm	new	ecigs	vaporizer	flavor
Ward	blufreedom	blufreedom	thanks	day	vapor	ecig	ecig
New	blu	blu	customer	save	liquid	vaping	save
Ecigs	rewards	vapelife	blunation	flavor	blog	ecigs	liquid
electronic	freedom	vaporlounge	call	products	categories	vape	vaping
Day	ecigs	indycar	sxsw	sale	shirt	vapesess	kit
Cigs	rewards	electriclounge	new	cig	standard	flavor	sale
Org	freedom	daysofblu	help	kit	mens	sale	day

Table 4. Cross-site users.

	Facebook	Twitter	Instagram	Google+
Facebook	18504	32	60	0
Twitter	--	2048	128	0
Instagram	--	--	3613	0
Google+	--	--	--	266

Table 5. Non-content interactions between users and e-cigarette brands. The numbers in parentheses represent the ratio of comments to original posts on Facebook and Instagram.

Facebook	Count	Google+	Count	Instagram	Count	Twitter	Count
comments	26612			comments	2127		
Blu	18582 (2.94)			Blu	1838 (5.37)		
V2	8030 (2.23)			V2	289 (0.98)		
Likes	54029	plusoners	754	likes	17034		
Blu	37231	Blu	740	Blu	11214		
V2	16798	v2	14	V2	5820		
		resharers	80			retweets	14781
		Blu	80			Blu	13879
		V2	0			V2	902

Discussion

Exploring the Results

The 1-case data provides valuable information in determining some of the most discriminating terms used by each e-cigarette brand, for the different social networking sites. It also reveals early evidence of how the two e-cigarette brands differ in their social media marketing strategies, with V2 focusing on marketing products on their website and Blu using Twitter for user interactions and Facebook for political activity information. The 163 terms were able to show that specific combinations, specifically Blu/Twitter, Blu/Facebook, and V2/Twitter, were being utilized by each brand for a particular type of marketing. However, the limited data—only 163 terms—does not provide enough information for deeper investigations. Therefore, we viewed the data through several other lenses. First, we examined the n-case data to include all possible terms (Table 2). In many of these cases, there were technologically driven explanations for the high discriminating power of some of the terms. For example, *RT* has the highest TF-IDF score and is found in only two case scenarios (Blu/Twitter and V2/Twitter). However, this is not unexpected, as it is only used in Twitter as shorthand for "retweet". Similarly, terms such as *href* or *nofollow* are webpage-coding syntax used only in Google+ to create links to external sites, usually to images or videos. Interestingly, Instagram is the only site where coding terms are not used, another artifact of the available technology. As an image-based platform, there would be no need for text to contain any additional image links for each Instagram post. Other types of links seen in Facebook or Google+ were not seen in Instagram captions. Overall, the n-case results suggest that properties inherent to specific platforms can dictate the types of discriminating content found.

We continued to broaden our view by removing the discriminating IDF and also filtering out technology-specific terms (Table 3). These results help provide additional evidence of the marketing strategies of each e-cigarette brand on the different sites. When looking at the terms for V2, we found a similar theme as in the 1-case results: V2's focus is on brand marketing, different products, and directing users to their website. In the case of Blu, we again confirmed some of the findings from the 1-case view. Terms such as *thanks*, *customer*, and *help* in Twitter were indicative of interactions with their followers. The only political term, *ward* (in context of geopolitical boundaries), was found only in Facebook. Interestingly, we saw a theme across Blu's social networking site presence, containing the terms *blunation* and *blufreedom* in Google+, Instagram, and Twitter. Blu's social media strategy appears to focus more on community and lifestyle, contrasting sharply with V2's efforts to market products and direct users to their website. Blu's efforts are more engaging than V2, possibly leading to more conversations and additional activities across the sites. Table 5 shows evidence of user engagement; user comments represent actions when users are responding in conversation to an original post, as compared to likes or redirects, which typically only require a single mouse click. In both sites where comments were recorded (Facebook and Instagram), Blu followers commented at higher rates than V2.

The cross-platform network analysis found several meaningful results (Table 4). First, the users on Google+ appeared to share no connection with any of the other sites. One possible reason for this is that Google+ does not explicitly require a traditional screen name that serves as an alias, but instead concatenates what the user inputs for a first and last name. Because extra steps are needed to create a pseudonym, users are more likely to retain Google+'s universal naming scheme, resulting in different aliases used than on other sites. Second, the highest number of overlapping screen names was between Instagram and Twitter. We had expected Facebook to be one of the overlapping pairs based on the high number of collected Facebook users and Pew's survey of users with multiple social media accounts [20], leading to a much greater potential for matches. However, as Instagram and Twitter have a more distinct focus on their media usage—short text for Twitter, images for Instagram—it is possible that users would overlap in their usage of these two sites rather than the more general social networking tools of Facebook. Overall, we expect that that our results are likely to represent the minimum amount of overlapping users, as it is likely that users might maintain different online identities and change their screen names in different sites.

The content analysis of cross-platform posts yielded no results where activity from a user on one site led to sharing that

message on a different site. While our methods were sound, we found several limitations to consider. When we searched for identical content between Instagram and another site, either Twitter or Facebook, we found that an image-based post would be difficult to forward to another site; the user would need to save the image and repost it, otherwise the caption text would not make sense out of context. Conversely, a message on Twitter without an image would not be posted on Instagram. Also, it is important to consider that the nature of one's network of friends on each site might be distinct. Different sites might be used to develop and maintain different relationships [16]. People might be using Facebook to connect with close family and friends who may not be supportive of smoking/e-cigarette use, therefore making Facebook an unbefitting platform to post about e-cigarettes. Twitter, on the other hand, allows users to connect directly with the staff and management of e-cigarette companies, which makes it easier for them to communicate and seek support directly through tweets. This is also apparent from the increased focus of the companies on Twitter use (V2=94%, Blu=66%). Another limitation was in attempting to study any content posted on Facebook, as most users had made their accounts private.

While not directly related to our research questions, we were also able to observe some of the consequences of each brand's marketing strategy. We compared the available interactions that users could have with each e-cigarette brand for the four websites; these included Facebook's comments and likes, Google+'s plusoners and resharers, Instagram's comments and likes, and Twitter's retweets. In every case, Blu's followers always had higher percentages of interactions (see Table 5). This suggests that Blu's efforts at interacting with their user base, rather than V2's strategy of directing traffic to their website, is more successful at engaging users, eliciting responses, and raising interest. These strategies might be based on the differences between an independent brand and one owned by a tobacco company. Additional studies with other brands will be necessary to determine if parent-company ownership has any effect.

Addressing the Research Questions

In addressing the first research question—Are e-cigarette brands exploiting the affordances of each site in their marketing?—we found evidence that Blu and V2 were using the sites in different ways, likely utilizing the affordances made available. The methods by which each brand interacted with its audience, advertised products, or relayed information aligned with the affordances of each platform. We could see Blu's usage of Twitter as a medium to interact with its followers, frequently mentioning users by name, and directly conversing with them. Twitter's @user_mention function, and an interface that allows users to immediately see Blu's interactions with them, all support an interactive environment. Facebook's "wall" feature offers a different format, in which users immediately see all other posts connected to a single parent discussion message. This affords a system whereby mass broadcasting is effective, reflected in the high number of political announcements and news events posted, with almost no direct user interaction. Kietzmann et al's [15] view of Facebook's affordances reflects this, as they note the importance of relationships over conversations. Twitter, on the other hand, prioritizes conversations over relationships, supporting Blu's usage. V2 similarly utilized Twitter functions for easy interaction with individual users, although they consistently used all of the sites as a way to link users back to their homepage.

In addressing the second research question—Are e-cigarette brands targeting sub-populations (eg, women, teenagers) by taking advantage of the demographic differences of different social networking sites?—we found no data suggesting direct targeting of sub-populations based on social networking site. The results of the 1-case view found it was likely that some of the sites were being used in different ways, but not in a manner consistent with focusing on any specific demographics. We followed up by analyzing the data from each site during specific holidays that might show a favoring of announcements or advertisements for one site over another and combined this information with the trends in different demographics reported by Pew on each of the sites [20]. For example, we examined the content during Mother's Day, Father's Day, Cinco de Mayo, and Black History Month, but found that the companies either ignored the holiday or were consistent in their celebration notices across all sites.

In addressing the third research question—Are there identifiable instances of users and information crossing different social networking sites?—we found a small percentage of users that had identical screen names on multiple sites, with the highest percentage being from Instagram/Twitter (128). These numbers differed from Pew's report that multi-site users tended to use Facebook with another platform [20], although the numbers are too small to provide real contradictory evidence, and our study used a conservative comparison of screen names and profiles. We found no direct results of users acting on content by Blu or V2 leading to posting the same content on another site. We also discovered that matching user activity across different platforms can be problematic, as different types of activity were collected based on what is available from each site: Facebook's posts, comments, and likes; Twitter's retweets, Instagram's comments and likes; and Google+'s resharers and plusoners (Table 5). Our findings will greatly support future research in multi-social networking site studies, as we can build on how to find users, analyze content, and normalize different activities.

Conclusion

The purpose of this study was to examine the marketing strategies of leading e-cigarette brands on multiple social networking sites. We found that the two brands, Blu and V2, utilize different methods in how they interact with their customers. While Blu tends to be more conversational, V2's focus is to direct people to their website. We did not find any evidence that either brand targets vulnerable populations in their strategies. However, we are able to see that Blu harnesses the affordances of different sites to conduct different types of marketing to their customer base. Our results demonstrate that a content analysis of multiple social networking sites can serve to identify how companies or brands can vary their marketing strategies based on the available technologies. We plan on expanding the foundation of this methodology to further

understand potential diffusion of information across multiple sites.

There are several limitations in this study. Primarily, we focused on only two e-cigarette brands. While their dominant online presence made them attractive candidates, we are also interested in following up with additional brands. Similarly, we only chose four social networking sites. Other social networking sites such as Pinterest, Tumblr, and LinkedIn were considered but were not included due to various limitations. We also did not interview any of the users in the study for specific reactions to or comments on the messages by the e-cigarette brands. Presenting the user side and experience would help provide additional context to the media messages. However, two of the authors are involved in a parallel project that, informed by some of the results of this pilot, will involve such interviews. Relatedly, we limited our content to publicly available data. While this affords easy access since all of the text or images posted can be viewed by anyone online, we do not know the impact on the larger online audience, which includes users that restrict access to their accounts. Therefore, the ability to generalize the results should be limited to public user accounts. Lastly, our content analysis focused on individual terms and we did not develop a synonym set of conceptually equivalent themes prior to the analysis. While this allows for a cleaner analysis, it risks over-representation of certain concepts.

We strongly believe cross-platform studies will be a vital area of research in the future. There are apps available that allow users to streamline their post activity by broadcasting content, whether text, image, or other media, to multiple social networking sites at once (eg, Everypost, HootSuite). New social networking sites are also continuing to be developed, especially as technology provides new opportunities in other hardware, such as mobile phones or wearable devices. A fast-paced arena makes single-platform studies very narrow and difficult to generalize, as more of the population continues using multiple social networking sites [20]. Additionally, public health campaigns have begun utilizing social networking sites as a platform for their causes, although not always with the desired outcomes [24]. As researchers, we must be able to adapt to the changing landscape of social media tools. In particular, public health researchers should be made aware what sites might be susceptible to potentially dangerous marketing strategies. Our research in multi-social networking site marketing should only be considered a first step in understanding this type of analysis. New tools in social media will require constant vigilance by those in public health to not only understand new arenas, but also to quickly develop strategies for prevention and intervention. Researchers need to understand what tools are available to establish counter-measures, whether against demographic targeting or misinformation. Public health campaigns should be carried out across the social networking sites that will best reach influential users able to spread and diffuse messages across many sites. Researchers must know what affordances need to be focused on and how users are affected, rather than blindly choose their platforms and audience.

Acknowledgments

Research reported in this publication was supported by grant number P50CA180905 from the National Cancer Institute and FDA Center for Tobacco Products.

Authors' Contributions

KC was the primary author and contributed to the planning, study design, analyses, and manuscript. AS and TV contributed to the analyses and manuscript.

Conflicts of Interest

None declared.

References

1. FDA, ” 2. Available Online. “Public Health Focus - Electronic Cigarettes (e-Cigarettes)" URL: http://www.fda.gov/newsevents/publichealthfocus/ucm172906.htm [accessed 2015-09-04] [WebCite Cache ID 6bJ5uDIVx]
2. Pearson JL, Richardson A, Niaura RS, Vallone DM, Abrams DB. e-Cigarette awareness, use, and harm perceptions in US adults. Am J Public Health 2012 Sep;102(9):1758-1766 [FREE Full text] [doi: 10.2105/AJPH.2011.300526] [Medline: 22813087]
3. King BA, Alam S, Promoff G, Arrazola R, Dube SR. Awareness and ever-use of electronic cigarettes among U.S. adults, 2010-2011. Nicotine Tob Res 2013 Sep;15(9):1623-1627. [doi: 10.1093/ntr/ntt013] [Medline: 23449421]
4. Regan AK, Promoff G, Dube SR, Arrazola R. Electronic nicotine delivery systems: adult use and awareness of the 'e-cigarette' in the USA. Tob Control 2013 Jan;22(1):19-23. [doi: 10.1136/tobaccocontrol-2011-050044] [Medline: 22034071]
5. Richardson A, Ganz O, Stalgaitis C, Abrams D, Vallone D. Noncombustible tobacco product advertising: how companies are selling the new face of tobacco. Nicotine Tob Res 2014 May;16(5):606-614. [doi: 10.1093/ntr/ntt200] [Medline: 24379146]
6. Federal Trade Commission. Federal Trade Commission Cigarette Report for 2011. Washington: Federal Trade Commission; 2013 May. Cigarette Report for 2011 URL: https://www.ftc.gov/sites/default/files/documents/reports/

federal-trade-commission-cigarette-report-2011/130521cigarettereport.pdf [accessed 2015-09-06] [WebCite Cache ID 6bM82BcbL]
7. Bullen C, Howe C, Laugesen M, McRobbie H, Parag V, Williman J, et al. Electronic cigarettes for smoking cessation: a randomised controlled trial. Lancet 2013 Nov 16;382(9905):1629-1637. [doi: 10.1016/S0140-6736(13)61842-5] [Medline: 24029165]
8. Caponnetto P, Campagna D, Cibella F, Morjaria JB, Caruso M, Russo C, "EffiCiencySafety of an eLectronic cigAreTte (ECLAT) as Tobacco Cigarettes Substitute: A Prospective 12-Month Randomized Control Design Study, et al. vol. 8, no. 6,, Jun 2013:e66317. [doi: 10.1371/journal.pone.0066317]
9. Grana RA, Ling PM. "Smoking revolution": a content analysis of electronic cigarette retail websites. Am J Prev Med 2014 Apr;46(4):395-403 [FREE Full text] [doi: 10.1016/j.amepre.2013.12.010] [Medline: 24650842]
10. Etter JF, Bullen C. A longitudinal study of electronic cigarette users. Addict Behav 2014 Feb;39(2):491-494. [doi: 10.1016/j.addbeh.2013.10.028] [Medline: 24229843]
11. Dawkins L, Turner J, Hasna S, Soar K. The electronic-cigarette: effects on desire to smoke, withdrawal symptoms and cognition. Addict Behav 2012 Aug;37(8):970-973. [doi: 10.1016/j.addbeh.2012.03.004] [Medline: 22503574]
12. Huang J, Kornfield R, Szczypka G, Emery SL. A cross-sectional examination of marketing of electronic cigarettes on Twitter. Tob Control 2014 Jul;23 Suppl 3:iii26-iii30 [FREE Full text] [doi: 10.1136/tobaccocontrol-2014-051551] [Medline: 24935894]
13. Gibson J, " IP, Acting, Knowing R. "The theory of affordances". S. & J. Bransford, Ed. Hillsdale, NJ: Lawrence Erlbaum Associations 1977.
14. Norman D. The psychology of everyday things. New York: Doubleday. ISBN: 0465067107 1988.
15. Kietzmann JH, Hermkens K, McCarthy IP, Silvestre BS. "Social media? Get serious! Understanding the functional building blocks of social media," Bus. Horiz., vol. 54, no. 3, pp. 241-251, May :2011. [doi: 10.1016/j.bushor.2011.01.005]
16. Chu KH, Suthers DD. Technology-Mediated Relationships in a Socio-technical System. In: Proceedings of the 46th Hawaii International Conference on System Sciences.: IEEE; 2013 Presented at: 2013 46th Hawaii International Conference on System Sciences (HICSS); January 7-10, 2013; Wailea, Maui, HI URL: http://ieeexplore.ieee.org/xpl/articleDetails.jsp?arnumber=6479926 [doi: 10.1109/HICSS.2013.511]
17. Duggan M, Smith A. Social Media Update 2013.: Pew Internet and American Life Project; 2013 Dec 30. Social Media Update 2013 URL: http://www.pewinternet.org/2013/12/30/social-media-update-2013/ [accessed 2015-09-06] [WebCite Cache ID 6bM8XFWnG]
18. Mislove A, Lehmann S, Ahn YY, Onnela JP, Rosenquist JN. Understanding the Demographics of Twitter Users. In: Proceedings of the Fifth Int. AAAI Conf. Weblogs Soc. Media.: AAAI; 2011 Presented at: Fifth Int. AAAI Conf. Weblogs Soc. Media; July 17-21, 2011; Barcelona, Spain URL: http://dougleschan.com/the-recruitment-guru/wp-content/uploads/2014/01/Understanding-the-Demographics-of-Twitter-Users-Jukka-Pekka-....pdf
19. Chang J, Rosenn I, Backstrom L, Marlow C. ePluribus: Ethnicity on Social Networks. In: Proceedings of the Fourth Int. AAAI Conf. Weblogs Soc. Media.: AAAI Press; 2010 Presented at: Fourth Int. AAAI Conf. Weblogs Soc. Media; May 23-26, 2010; Washington, DC URL: http://www.aaai.org/ocs/index.php/ICWSM/ICWSM10/paper/viewFile/1534/1828
20. Duggan M, Ellison N, Lampe C, Lenhart A, Madden M. "Social Media Update 2014," Pew Research Center's Internet & American Life Project. 2014. URL: http://www.pewinternet.org/2015/01/09/social-media-update-2014/ [accessed 2015-01-15] [WebCite Cache ID 6VbRMf5n5]
21. Madden M, Lenhart A, Cortesi S, Gasser U, Duggan M, Smith A. "Teens, Social Media, and Privacy,". 2013. URL: http://www.pewinternet.org/2013/05/21/teens-social-media-and-privacy/ [accessed 2014-03-15] [WebCite Cache ID 6O5psgdAN]
22. Kim AE, Arnold KY, Makarenko O. E-cigarette advertising expenditures in the U.S., 2011-2012. Am J Prev Med 2014 Apr;46(4):409-412. [doi: 10.1016/j.amepre.2013.11.003] [Medline: 24650844]
23. Salton G, Fox EA, Wu H, "Extended Boolean Information Retrieval, " CACM. vol. 26, no. 11, pp. 1022-1036, Nov 1983. [doi: 10.1145/182.358466]
24. Harris J, Moreland-Russell S, Choucair B, Mansour R, Staub M, Simmons K. Tweeting for and Against Public Health Policy: Response to the Chicago Department of Public Health's Electronic Cigarette Twitter Campaign. J Med Internet Res 2014;16(10):e238 [FREE Full text] [doi: 10.2196/jmir.3622]

Edited by G Eysenbach; submitted 29.05.15; peer-reviewed by A Richardson, A Sanders-Jackson, S Leischow, B Liang; comments to author 09.07.15; revised version received 07.08.15; accepted 17.08.15; published 11.09.15.

Please cite as:
Chu KH, Sidhu AK, Valente TW
Electronic Cigarette Marketing Online: a Multi-Site, Multi-Product Comparison
JMIR Public Health Surveill 2015;1(2):e11
URL: http://publichealth.jmir.org/2015/2/e11/
doi:10.2196/publichealth.4777
PMID:27227129

Original Paper

The Effect of Commuting Patterns on HIV Care Attendance Among Men Who Have Sex With Men (MSM) in Atlanta, Georgia

Sharoda Dasgupta[1], PhD, MPH; Michael R Kramer[1], PhD; Eli S Rosenberg[1], PhD; Travis H Sanchez[1], DVM, MPH; Landon Reed[2], MCRP, MS; Patrick S Sullivan[1], DVM, PhD

[1]Emory University, Atlanta, GA, United States

[2]Atlanta Regional Commission, Atlanta, GA, United States

Corresponding Author:
Sharoda Dasgupta, PhD, MPH
Emory University
1518 Clifton Road NE
Atlanta, GA, 30322
United States
Phone: 1 5124841667
Fax: 1 8663118234
Email: sdasgu4@emory.edu

Abstract

Background: Travel-related barriers to human immunodeficiency virus (HIV) care, such as commute time and mode of transportation, have been reported in the United States.

Objective: The objective of the study was to investigate the association between public transportation use and HIV care attendance among a convenience sample of Atlanta-based, HIV-positive men who have sex with men (MSM), evaluate differences across regions of residence, and estimate the relationship between travel distance and time by mode of transportation taken to attend appointments.

Methods: We used Poisson regression to estimate the association between use of public transportation to attend HIV-related medical visits and frequency of care attendance over the previous 12 months. The relationship between travel distance and commute time was estimated using linear regression. Kriging was used to interpolate commute time to visually examine geographic differences in commuting patterns in relation to access to public transportation and population-based estimates of household vehicle ownership.

Results: Using public transportation was associated with lower rates of HIV care attendance compared to using private transportation, but only in south Atlanta (south: aRR: 0.75, 95% CI 0.56, 1.0, north: aRR: 0.90, 95% CI 0.71, 1.1). Participants living in south Atlanta were more likely to have longer commute times associated with attending HIV visits, have greater access to public transportation, and may live in areas with low vehicle ownership. A majority of attended HIV providers were located in north and central Atlanta, despite there being participants living all across the city. Estimated commute times per mile traveled were three times as high among public transit users compared to private transportation users.

Conclusions: Improving local public transit and implementing use of mobile clinics could help address travel-related barriers to HIV care.

(JMIR Public Health Surveill 2015;1(2):e10) doi:10.2196/publichealth.4525

KEYWORDS

men who have sex with men; HIV care; commuting patterns; public transportation

Introduction

Importance of Regular Human Immunodeficiency Virus Care Engagement

Men who have sex with men (MSM) accounted for 67% of all new human immunodeficiency virus (HIV) infections in the United States in 2012, despite accounting for only 3% of the population [1,2]. Regular medical care utilization among HIV patients is important in maintaining viral load suppression, reducing transmission to others [3-5], and improving survival over time [5,6]. Among newly diagnosed cases in Georgia in 2011, an estimated 46% were regularly engaged in HIV care and 45% had achieved viral suppression within 15 months of diagnosis [7]. Similar estimates in HIV care engagement were observed among a large cohort of MSM in Atlanta [6].

Transportation Factors as Barriers to Human Immunodeficiency Virus Care

Transportation-related factors, such as travel distance and commute time, have been reported as substantial barriers to general medical care and attending HIV appointments [8-15]. Transportation assistance was reported as an unmet need among 16% of those who needed it in a cross-sectional, nationally representative study of HIV-positive individuals engaged in care [16]. Travel distance and mode are both important predictors of commute time [17] and can influence travel times differentially by neighborhood, depending on availability of public transportation, household vehicle ownership, and traffic congestion patterns [18,19].

Compared to traveling by car, using public transportation is often associated with longer commute times and reduced convenience and flexibility in travel [17,20]. This is especially important in cities like Atlanta with limited public transportation and a strong dependence on travel by car [21]. Longer commute times can be a deterrent to attending care visits, especially with competing household and job responsibilities [8]. In Atlanta, the Metropolitan Atlanta Rapid Transit Authority (MARTA) has been the primary source of public transportation infrastructure for both bus and rail since the 1970s. MARTA, as well as other transit systems in the metro area, serves mostly urban areas in the city [22,23].

Atlanta has historically been a highly segregated city with respect to race and income [24], with differential access to public transportation. A general north-south pattern exists; with predominantly lower income, black neighborhoods in south Atlanta and mostly white neighborhoods in north Atlanta. Neighborhood contextual factors, such as availability of resources and socioeconomic deprivation, have been shown to be associated with negative physical and mental health outcomes [25-27]. Our study objectives were two-fold. First, we investigated whether public transportation use is associated with HIV care attendance and whether the association varies by region of residence (north vs south) in Atlanta. Second, because taking public transportation can strongly affect commute times, we also estimated the relationship between travel distance and commute time by mode of transportation. Identifying areas where travel-related factors might be barriers to HIV care can be beneficial in planning targeted structural interventions to improve health care utilization.

Methods

Study Methodology

Recruitment

The Engage Study, a cross-sectional study of self-identifying HIV-positive MSM, was designed to investigate structural and psychosocial barriers to HIV care among MSM living in the Atlanta area. A convenience sample of men was recruited from October 2012 to June 2013 from two sources: (1) based on participation in previously conducted Atlanta-based studies on HIV, and (2) from Facebook. Men who previously participated in the Atlanta-based studies and had a known positive HIV test were contacted by phone and email for recruitment. Individuals interested in participation were then sent a Web link to the Internet eligibility screener by email. Participants from Facebook were recruited based on banner advertisements targeting men who were interested in other men and lived within 50 miles of Atlanta. Those who clicked on the banner advertisements were directed to the Internet eligibility screener.

Individuals were eligible for participation if they reported being 18 years of age or older, being told they were HIV-positive by a health care provider, having sex with at least one other man in their lifetime, and living in the Atlanta area. All consenting participants were directed from the Internet eligibility screener to the questionnaire administered using a Health Insurance Portability and Accountability Act (HIPAA)-compliant Internet survey software platform, SurveyGizmo (Boulder, CO) [28]. The Emory University Institutional Review Board approved the study protocol (approval number: IRB00060430). More details on study methodology have been previously described [29].

Measures

The Web-based questionnaire collected information on demographic characteristics, potential structural (eg, public transportation use, health insurance status) and psychosocial barriers (eg, perceived HIV-related stigma, disclosure of HIV status, self-perceived community perceptions of HIV) to HIV care engagement, and characteristics related to HIV care experiences (eg, number of attended care appointments with the most recently attended HIV provider). We also collected information about home address at the time of the interview and location of the last HIV care provider, which was used to estimate travel distance and commute time to attend care visits. We geocoded all addresses using ArcGIS 10.2 (Redlands, CA). Participants were also asked about mode of transportation used to regularly attend care. Those reporting normally traveling by train, bus, or foot were considered to be public transportation users; otherwise, they were considered private transportation users.

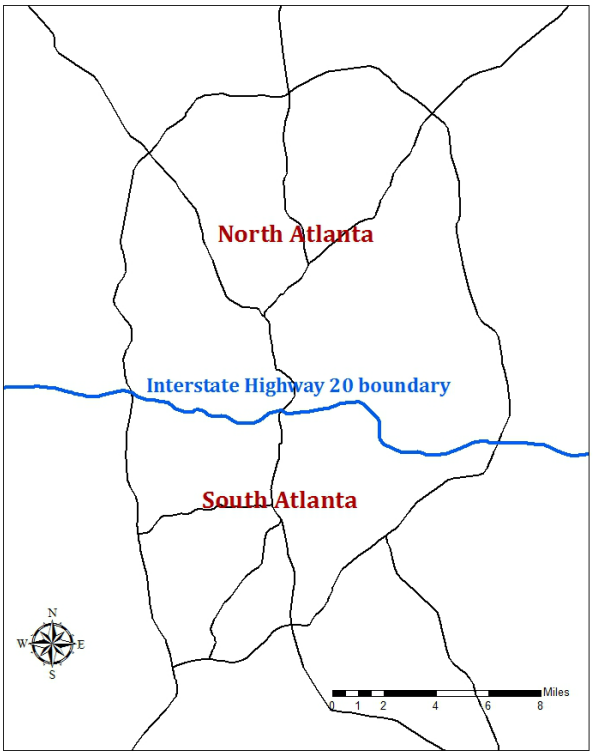
To estimate travel distance and commute time between participant residence and last attended HIV provider, we used the Google Maps Directions application programming interface. Distance and commute time were calculated for each pair of

origin-destination points (ie, residence and provider locations) based on the most optimal route chosen by Google maps. Travel parameters were calculated separately for those who took public transportation versus those who did not. Those who did not take public transportation were assumed to travel by car. All travel parameters were calculated assuming a departure day and time of Friday, March 7, 2014 at 10:00 AM. Latitude-longitude coordinates for residence were anonymized before entered into Google Maps to protect confidentiality of participants.

Since our research objectives focused on assessing differences in effect estimates across region of residence, we stratified the results by residence in north versus south Atlanta. Interstate highway 20 served as a coarse boundary for these regions, as it is often used to distinguish between areas of differing socioeconomic status, such as racial composition and average household income [24] (Figure 1 shows this).

We obtained information on household vehicle access, which we used as a proxy for vehicle ownership, from the US Census Bureau [30] and data on availability of public transportation bus and train routes from the Atlanta Regional Commission [31].

Figure 1. Geographic boundaries defining regions of residence used in analyses.



Analytic Methods

Descriptive Statistics

Descriptive statistics were computed for transportation-related factors, demographic characteristics, and the number of attended HIV care appointments in the previous 12 months, overall and by region of residence (north vs south). We reported medians and interquartile ranges (IQR) for continuous variables, and counts and frequencies for categorical variables. Wilcoxon-Mann-Whitney tests were used to evaluate differences in continuous variables across region of residence. For categorical variables, differences were assessed using the Mantel-Haenszel chi-square test.

Modeling

Public Transportation and Human Immunodeficiency Virus Care Attendance

A Poisson regression model estimated the rate of attended HIV care-related visits with the most recent provider in the past 12 months (using attended HIV care appointment counts) as a function of whether or not public transportation was taken to attend care, and examined whether the association was modified by region of residence in the city. The offset variable represented eligible days to receive HIV care in the past 12 months, and was coded as the natural log of 365 days unless date of diagnosis was less than a year before the survey was completed, in which case it represented the number of days between the date of the survey and date of diagnosis.

Race was considered an important confounding variable, and thus was included, *a priori*, in the final, multivariable model. Mode of transportation, region of residence, and the interaction between the two variables were also retained in the final model, since they were the primary explanatory variables of interest. For other covariates, bivariate associations with the outcome of interest were assessed, and variables with *P* values of less than 0.1 were eligible for possible inclusion in the final multivariable model. Except for the variables included *a priori* in the analysis, backward selection was used to determine which

variables should be retained in the final model, using a cutoff of $P<.05$. The multivariable Poisson model was built using SAS 9.3 (Cary, NC).

In a post-hoc analysis, we examined spatial relationships between estimated commute time, the network of available public transportation routes in Atlanta, and areas with low household vehicle ownership. Using ArcGIS 10.2, we utilized kriging to interpolate commute time associated with traveling to the last attended HIV provider. Because we did not ask participants directly about individual household vehicle ownership, we used US Census data as a marker for study areas with poor access to a vehicle. We defined vehicle ownership as having access to one or more vehicles in the household (a proxy for ownership); census tracts with < 87% household vehicle access were considered areas of low ownership. We based this cutoff on results from the 2009 National Household Travel Survey, which estimated that approximately 13% of US households were without access to a vehicle in large urban areas [32]. Using GeoDa (Tempe, AZ), local Moran's I statistics with significance testing (alpha = .05) evaluated local spatial autocorrelation to identify clusters of low vehicle ownership geographically.

Travel Distance and Commute Time

We used a linear regression model to describe the relationship between travel distance and commute time, stratified by mode of transportation used to attend appointments. For each mode of travel, the intercept represented initial investment in time; the slope provided information on the increase in commute time for each mile traveled. No other covariates were of interest, and therefore, were not included in the final model.

Results

Descriptive Statistics

A total of 213 eligible MSM participated; 205/213 (96.2%) participants reported ever receiving HIV care, among which 184/205 (89.7%) reported valid location data on home address and last HIV provider to enable calculation of road distance and commute time between the two. A total of 178/184 (96.7%) respondents who traveled less than 100 miles and lived within 50 miles from the center of Atlanta were used in the final analysis dataset.

The median age of participants was 34 years old, over half of participants reported an annual household income of less than US $20,000, and about two-thirds identified as black/African American race (Table 1). Participants attended a median of 3 appointments with their most recent HIV care provider in the previous 12 months, and about a third of participants reported missing at least one appointment. Overall, (72/178) 40% reported using some form of public transportation to attend care; median commute time was 22 minutes and median travel distance was about 9 miles.

Participants living in south Atlanta were significantly more likely to report black race ($P<.001$), have lower annual household income ($P=.04$), and not have health insurance at the time of the survey ($P=.03$). Greater reported use of public transportation ($P=.002$), travel distance ($P=.003$), and commute times ($P<.001$) associated with attending HIV care visits were observed in south Atlanta, compared to north Atlanta. Participants in south Atlanta were also more likely to live in census tracts with low vehicle ownership, but this difference was not statistically significant ($P=.05$).

Table 1. Demographic characteristics reported among a convenience sample of HIV-positive MSM linked to care in Atlanta, Georgia, 2012-2013.

Demographic characteristics		Overall[a]			North Atlanta[a]			South Atlanta[a]		
		n	%	Mean visits	n	%	Mean visits	n	%	Mean visits
People living in tracts with low household vehicle ownership										
	< 87%	64	36	3.5	41	33	3.5	23	44	3.3
	> 87%	114	64	3.2	85	67	3.1	29	56	3.6
Taking public transit (bus, train, foot) to attend care visits [b]										
	Yes	72	40	3.4	42	33	3.4	30	58	3.3
	No	106	60	3.3	84	67	3.1	22	42	3.8
Age in years										
	< 35 years	91	51	2.9	65	52	3.0	26	50	2.7
	> 35 years	87	49	3.7	61	48	3.4	26	50	4.3
Race [b]										
	White	60	34	3.3	52	41	3.1	8	15	4.5
	Black/African American	107	60	3.4	66	52	3.4	41	79	3.3
Education										
	High school or less	32	18	3.2	20	16	2.9	12	23	3.6
	At least some college	144	81	3.4	104	83	3.3	40	77	3.5
Annual household income [b] **(US)**										
	< $20,000	93	52	3.3	61	48	3.2	32	62	3.6
	> $20,000	79	44	3.2	62	49	3.1	17	33	3.6
Current health insurance status [b]										
	Yes	102	57	3.3	78	62	3.2	24	46	3.6
	No	74	42	3.3	46	37	3.2	28	54	3.4

[a] Numbers may not sum to total because of missing values. Percentages may not add up to 100 due to rounding.

[b] Statistically significant differences observed across region of residence, alpha =.05.

Modeling

Public Transportation and Human Immunodeficiency Virus Care Attendance

Of those living south Atlanta, the adjusted rate of HIV care attendance was 25% lower among those who took public transportation to attend care visits, compared to those who took private transportation (aRR: 0.75, 95% CI 0.56, 1.0; Figure 2 shows this). No significant association was observed among those living in north Atlanta (aRR: 0.90, 95% CI 0.71, 1.1). The multivariable model adjusted for race, annual household income, and health insurance status reported at the time of the interview. Although the interaction between region of residence and use of public transportation was not significant, stratified results are presented because we hypothesize that factors related to socioeconomic status (SES), such as reasons for taking public transportation to attend visits, might vary depending on region of residence

To explore this hypothesis, we examined interpolated commute times across urban areas of Atlanta with respect to census tracts with low car ownership and the network of available public transportation in the city (Figure 3 shows this). Interpolated commute times greater than the overall mean (34 minutes) were observed in much of south Atlanta. A majority of attended HIV providers were also located in north or central Atlanta, which may have driven longer commute times observed for participants living in south Atlanta.

Further, positive spatial autocorrelation was observed in a cluster of census tracts with low car ownership in south Atlanta. Within the auxiliary interstate highway 285 (often used as a boundary for urban vs suburban/rural areas of Atlanta), there are 6.6 miles of available public transportation per 1000 population in south Atlanta and 4.4 miles of available transit per 1000 population in north Atlanta.

Figure 2. Adjusted association between public transportation use and rates of HIV care attendance in the past 12 months, by region of residence, among a convenience sample of HIV-positive MSM linked to care, Atlanta, Georgia, 2012-2013. Final multivariable model controls for race, annual household income, and health insurance status reported at the time of interview.

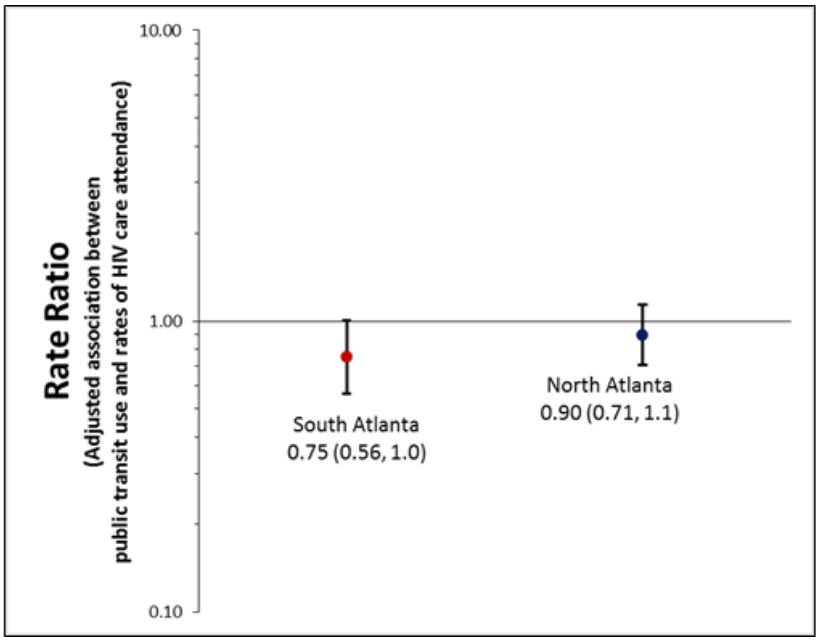
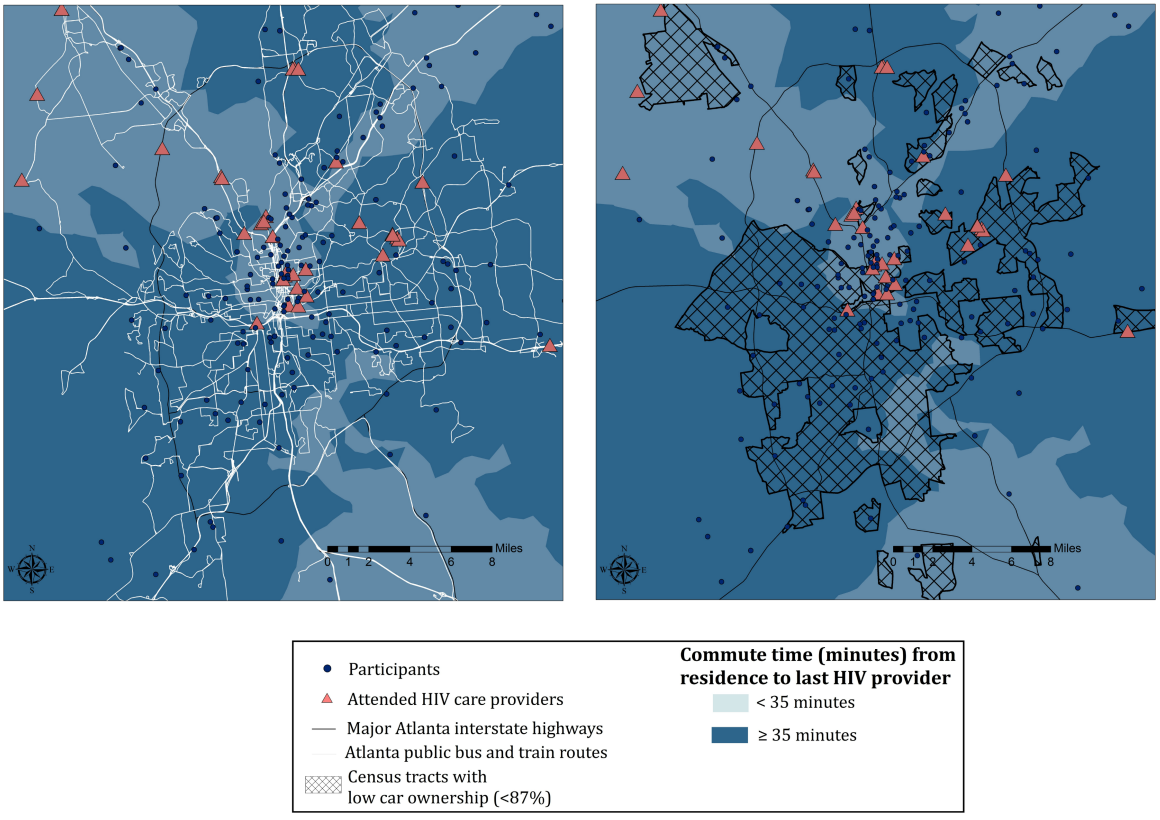


Figure 3. Relationships between interpolated commute time and network of public transportation routes (left), and interpolated commute time and low car ownership (right), among a convenience sample of HIV-positive MSM linked to care, Atlanta, Georgia, 2012-2013. Locations of HIV providers attended by participants are denoted in pink triangles, and locations of participant residences, which have been anonymized, are represented by dark blue dots.



Travel Distance and Commute Time

The modeled estimates showed that the relationship between travel distance and commute time varied by mode of transportation taken to attend HIV care visits (Table 2 and Figure 4 show this). The model explained 93% of the variance of the data around the estimated regression equation. The estimated initial time investment associated with commuting

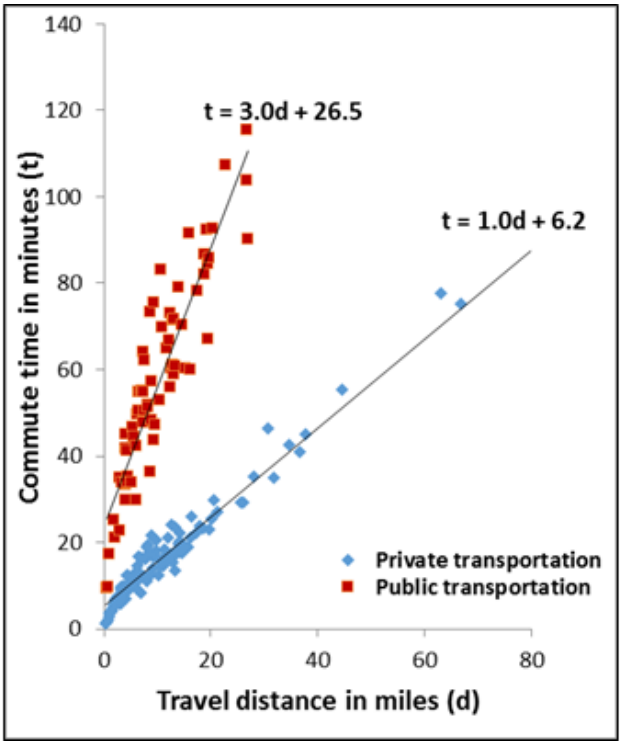
was over 4 times higher among public transportation users (27 minutes) compared to private transportation users (6 minutes). Among those who took private transportation, each mile of travel resulted in an additional minute of commute time; by contrast, the rate of increase in commute time per mile traveled was 3 times as high among those who took public transportation. Estimated commute times were consistently longer for public transportation users; differences in commute times for key distance values are provided in Table 2.

Table 2. Modeled estimates for initial time investment, rate of increase in commute time per mile traveled, and differences in overall commute time (for key distances) by mode of transportation taken to attend HIV care visits among a convenience sample of HIV-positive MSM linked to care in Atlanta, Georgia, 2012-2013.

Mode of transportation	Initial investment (minutes)	Δ Commute time per mile traveled (minutes)	Modeled commute time (minutes) for miles traveled			
			1 mile	5 miles	10 miles	15 miles
Public[a]	26.5 (23.6, 29.5)	3.0 (2.8, 3.3)	29.5	41.7	56.9	72.0
Private[a]	6.2 (4.5, 7.9)	1.0 (0.9, 1.1)	7.2	11.4	16.5	21.7

[a] Change in modeled commute times by mode of transportation for key travel distances, miles, listed in table: 22.3 minutes, 30.3 minutes, 40.3 minutes, 50.4 minutes.

Figure 4. Linear relationship between travel distance (miles) and commute time (minutes) to HIV care appointments, stratified by use of public transportation to attend HIV care appointments, among a convenience sample of HIV-positive MSM linked to care, Atlanta, Georgia, 2012-2013. In each of the estimated regression equations below, t represents commute time in minutes, d represents travel distance in miles, and the intercept represents the estimated time investment made in taking a specified mode of travel to attend HIV appointments.



Discussion

Principal Findings

In this study, we investigated commuting patterns related to attending HIV care visits, a topic which has not been extensively explored among HIV-positive MSM in Atlanta. Among those living in south Atlanta, using public transportation was associated with lower rates of HIV care attendance, compared to using private transportation. Participants in south Atlanta had greater access to public transportation (miles/1000 population), but traveled longer and further to attend HIV appointments and may be more likely to live in areas with low vehicle ownership. Both initial time investment and rate of increase in commute time per mile traveled to attend HIV visits were significantly higher among those who took public transportation, compared to those who did not.

Although not statistically significant, geographic differences in the association between public transportation use and care attendance could signify that transportation was more of a barrier to attending HIV care visits in south Atlanta. Because the sample size was limited, a larger study may have detected statistically significant differences in the association. We hypothesize that if the geographic differences in effect estimates exist, they may be driven by factors related to SES, such as differing reasons for taking public transportation, or differences in availability of medical resources around the metro area.

Although using public transportation is often associated with longer and more variable commute times and reduced flexibility in travel, there may be many reasons why public transportation is preferred, including: (1) concerns related to traffic congestion and pollution, (2) cost reductions associated with traveling by car, (3) convenience, if residence is in an urban area with access to public transit and limited space for private vehicle parking, and (4) not having another means of travel [17,20,33,34]. Out of these four reasons, the first three are related to convenience, or choice to take public transportation, while the fourth is associated with necessity because of lack of vehicle ownership. Because levels of household vehicle ownership may be higher in north versus south Atlanta, we hypothesize that those living in south Atlanta might be more likely to take public transportation out of necessity, and people living in north Atlanta might choose to take public transportation out of convenience. Although we did not directly measure reasons for taking certain modes of transit in the present study, exploring differences in reasons for taking public transportation in the future may be helpful in understanding complex patterns and dynamics between travel and medical care utilization.

When we examined a combination of population-based transportation-related factors with the Engage Study data to get a clearer picture of reasons for taking public transportation across Atlanta, we found that south Atlanta had overlapping geographic areas of longer estimated commute times to attend HIV care visits, low car ownership, and greater access to public transportation. Historically, south Atlanta also has a majority black population and greater levels of poverty compared to other areas of Atlanta [24], and along with downtown Atlanta, also has a greater burden of HIV compared to other areas of the city [35]. National data also show disproportionately greater use of public transportation among minorities and individuals from low-income households, suggesting socioeconomic differences in travel behaviors [36]. Therefore, transportation-related barriers to HIV care may be more prevalent in economically disadvantaged communities in south Atlanta where there is a greater need for HIV medical care utilization.

Although reasons for taking public transportation are highly correlated with SES, controlling for census tract-level car ownership and individual-level race and income in this analysis did not explain the observed association between public transportation use and HIV care attendance in south Atlanta. Individual-level household vehicle ownership could have explained the association if the participants are not representative of their census tract of residence. However, information on individual-level vehicle ownership was not available. Alternatively, there may have been one or more unmeasured factors associated with neighborhood economic disadvantage and deprivation, which explain the differential results between north and south Atlanta, and this should be further explored.

Differences in the density of available medical resources may have also helped drive the geographic differences in effect estimates. In particular, there were very few attended HIV providers located in south Atlanta, compared to north Atlanta. This is consistent with another study, which found poorer spatial accessibility to HIV providers in south Atlanta, where HIV prevalence is high [37]. Having fewer available providers in an area where residents are potentially more reliant on public transit as a sole means of travel might amplify travel-related barriers. Exploring this idea further in future studies through focus groups may help elucidate important drivers of travel-related barriers to HIV care.

Although this is a hypothesis-generating study with exploratory objectives, the results justify exploring in the future whether travel-related barriers affect medical care attendance differentially by region of residence among Atlanta-based, HIV-positive MSM. Larger studies which collect information on individual car ownership and any unmeasured factors which could potentially explain the differential effect estimates would help inform whether interventions related to improving spatial access might be beneficial. For instance, if transportation did indeed differentially affect HIV care attendance, the use of mobile clinics, as well as expansion of public transportation networks and more frequently operating bus and train routes, could be helpful in mitigating travel-related barriers.

Mobile vehicles used for HIV testing have been accepted by patients both in and out of the United States [38-40], but have rarely been used to administer HIV care, despite such an option being suggested to reduce transportation and socioeconomic barriers to medical care [41]. Mobile clinics have been used to provide other types of medical care previously, and have been associated with improved health care utilization [42], and potentially, fewer visits to the emergency department [43], after implementation.

Improving public transportation connectivity to other parts of the city, where preferred HIV clinics may operate, is key to increasing mobility of lower income communities that may be less likely to own a vehicle. Increasing frequency of existing public bus and train routes may also cut down on commute times and improve the level of convenience associated with taking public transportation. However, expansion of public transit in Atlanta has continually been a contentious issue among the public [23,44]. The original plan for a public transit system was published in a 1961 report by the Atlanta Region Metropolitan Planning Commission, and included an expansive, 66 mile rail network and covered five counties in the metro area [21,22]. Unfortunately, the plan was not approved by voters and eventually resulted in the 48 mile rail and 91 route bus system that it is today [22]. Despite limited availability of funds to expand the current public transportation network, incorporating discussions about public health during transit planning would be helpful in serving communities, which may benefit from greater access to medical resources they may need.

Limitations

There are several limitations to this study. First, the cross-sectional study design does not lend itself to making inferences on temporality or causality. Second, because a large proportion of participants were recruited on the Internet, HIV status was self-reported and could not be verified. However, we suspect little to no misclassification of HIV status because the study survey contained extensive questions about provider location and HIV care engagement. The primary outcome, number of attended HIV care visits in the past 12 months, was

self-reported, and therefore, is subject to information bias. We hypothesize that the number of attended appointments might be overreported, but do not suspect that misclassification was differential with respect to travel parameters.

The relationship between travel-related factors and HIV care attendance may be confounded by level of disease progression, which should be incorporated in future analyses. Results were generated from a convenience sample, which may not be representative of all HIV-positive MSM living in Atlanta, limiting generalizability of results. Obtaining more information on mode of transportation used to attend visits, including whether the patient carpooled with someone or received a ride and reasons for taking certain modes of transit, would have added to the results. In addition, using population-based estimates of vehicle access as a proxy for household vehicle ownership for participants may not have been appropriate due to convenience sampling. Finally, home and provider locations are based on self-report and are subject to information bias.

Conclusions

Using public transportation, compared to private transportation, may have been a barrier to HIV care among a sample of Atlanta-based, HIV-positive MSM living in south Atlanta. We hypothesize that reasons for taking public transportation and availability of HIV providers may differ across regions of residence in Atlanta, and, thus, could help explain the differences in the observed association by region. However, this hypothesis should be further explored in future studies.

The results from this analysis add to the current knowledge about travel and transportation-related barriers to HIV care, may inform the design of larger population-based studies which further explore potential neighborhood-level characteristics driving differences in travel-related barriers, and could provide guidance on potentially beneficial interventions which address gaps in care among Atlanta-based, HIV-positive MSM.

Acknowledgments

The National Institute of Mental Health (grant number R01MH085600), Eunice Kennedy Shriver National Institute for Child Health and Human Development (grant number R01HD067111), Emory University Center for AIDS Research (grant number P30AI050409), and Professional Development Support Funds provided by Laney Graduate School, Emory University financially supported the Engage Study.

Conflicts of Interest

None declared.

References

1. Centers for Disease Control and Prevention. Centers for Disease Control and Prevention. 2015. HIV surveillance report, 2012; Published 2014 URL: http://www.cdc.gov/hiv/library/reports/surveillance [accessed 2015-04-13] [WebCite Cache ID 6XlezndHa]
2. Purcell DW, Johnson CH, Lansky A, Prejean J, Stein R, Denning P, et al. Estimating the population size of men who have sex with men in the United States to obtain HIV and syphilis rates. Open AIDS J 2012;6:98-107 [FREE Full text] [doi: 10.2174/1874613601206010098] [Medline: 23049658]
3. Horstmann E, Brown J, Islam F, Buck J, Agins BD. Retaining HIV-infected patients in care: Where are we? Where do we go from here? Clin Infect Dis 2010 Mar 1;50(5):752-761 [FREE Full text] [doi: 10.1086/649933] [Medline: 20121413]
4. Mugavero MJ, Norton WE, Saag MS. Health care system and policy factors influencing engagement in HIV medical care: Piecing together the fragments of a fractured health care delivery system. Clin Infect Dis 2011 Jan 15;52 Suppl 2:S238-S246 [FREE Full text] [doi: 10.1093/cid/ciq048] [Medline: 21342913]
5. Das M, Chu PL, Santos G, Scheer S, Vittinghoff E, McFarland W, et al. Decreases in community viral load are accompanied by reductions in new HIV infections in San Francisco. PLoS One 2010;5(6):e11068 [FREE Full text] [doi: 10.1371/journal.pone.0011068] [Medline: 20548786]
6. 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]
7. HIV care continuum surveillance report, adults and adolescents, Georgia, 2012; Published 2014. on April 13, 2015 URL: https://dph.georgia.gov/sites/dph.georgia.gov/files/HIV%20Care%20Continuum%20Report_2012_web%20view_spreads.pdf [accessed 2015-04-13] [WebCite Cache ID 6Xlg7V4jg]
8. Cunningham WE, Andersen RM, Katz MH, Stein MD, Turner BJ, Crystal S, et al. The impact of competing subsistence needs and barriers on access to medical care for persons with human immunodeficiency virus receiving care in the United States. Med Care 1999 Dec;37(12):1270-1281. [Medline: 10599608]
9. Conviser R, Pounds MB. The role of ancillary services in client-centred systems of care. AIDS Care 2002 Aug;14 Suppl 1:S119-S131. [doi: 10.1080/09540120220150018] [Medline: 12204146]

10. Moneyham L, McLeod J, Boehme A, Wright L, Mugavero M, Seal P, et al. Perceived barriers to HIV care among HIV-infected women in the Deep South. J Assoc Nurses AIDS Care 2010;21(6):467-477. [doi: 10.1016/j.jana.2010.03.003] [Medline: 20430653]
11. Wohl AR, Carlos J, Tejero J, Dierst-Davies R, Daar ES, Khanlou H, et al. Barriers and unmet need for supportive services for HIV patients in care in Los Angeles County, California. AIDS Patient Care STDS 2011 Sep;25(9):525-532. [doi: 10.1089/apc.2011.0149] [Medline: 21774689]
12. Sagrestano LM, Clay J, Finerman R, Gooch J, Rapino M. Transportation vulnerability as a barrier to service utilization for HIV-positive individuals. AIDS Care 2014;26(3):314-319. [doi: 10.1080/09540121.2013.819403] [Medline: 23876086]
13. Buzza C, Ono SS, Turvey C, Wittrock S, Noble M, Reddy G, et al. Distance is relative: Unpacking a principal barrier in rural healthcare. J Gen Intern Med 2011 Nov;26 Suppl 2:648-654 [FREE Full text] [doi: 10.1007/s11606-011-1762-1] [Medline: 21989617]
14. Heckman TG, Somlai AM, Peters J, Walker J, Otto-Salaj L, Galdabini CA, et al. Barriers to care among persons living with HIV/AIDS in urban and rural areas. AIDS Care 1998 Jun;10(3):365-375. [doi: 10.1080/713612410] [Medline: 9828979]
15. Reif S, Golin CE, Smith SR. Barriers to accessing HIV/AIDS care in North Carolina: Rural and urban differences. AIDS Care 2005 Jul;17(5):558-565. [doi: 10.1080/09540120412331319750] [Medline: 16036242]
16. Blair JM, McNaghten AD, Frazier EL, Skarbinski J, Huang P, Heffelfinger JD. Clinical and behavioral characteristics of adults receiving medical care for HIV infection --- Medical Monitoring Project, United States, 2007. MMWR Surveill Summ 2011 Sep 2;60(11):1-20 [FREE Full text] [Medline: 21881551]
17. Beirão G, Sarsfield Cabral J. Understanding attitudes towards public transport and private car: A qualitative study. Transport Policy 2007 Nov;14(6):478-489. [doi: 10.1016/j.tranpol.2007.04.009]
18. Levinson D. Network structure and city size. PLoS One 2012;7(1):e29721 [FREE Full text] [doi: 10.1371/journal.pone.0029721] [Medline: 22253764]
19. Nurul Habib KM. Modeling commuting mode choice jointly with work start time and work duration. Transportation research part A: Policy and practice 2012 Jan;46(1):33-47. [doi: 10.1016/j.tra.2011.09.012]
20. Van Vugt M, Van Lange P, Meertens R. Commuting by car or public transportation? A social dilemma analysis of travel mode judgements. Eur. J. Soc. Psychol 1996 May;26(3):373-395. [doi: 10.1002/(SICI)1099-0992(199605)26:3<373::AID-EJSP760>3.0.CO;2-1]
21. University LB. ScholarWorks. Snow Jam 2014: Georgia State University,; 2014. Planning Atlanta, ryan gravel, and a car-dependent region URL: http://webapps.library.gsu.edu/blogs/library/2014/02/06/snow-jam-2014-planning-atlanta-ryan-gravel-and-a-car-dependent-region/ [accessed 2015-04-13] [WebCite Cache ID 6Xli6bjH0]
22. Metropolitan Atlanta Rapid Transportation Authority. Metropolitan Atlanta Rapid Transportation Authority. USA: MARTA; 2013. FY 2013 annual report: July 1, 2012-June 30, 2013 URL: http://jkheneghan.com/city/meetings/2013/Sep/FY2013_MARTOC_Report_Marta.pdf [accessed 2015-08-18] [WebCite Cache ID 6asao0Cuo]
23. Monroe D. Atlanta. Where it all went wrong: If only we could undo the MARTA Compromise of 1971; 2012. URL: http://www.atlantamagazine.com/great-reads/marta-tsplost-transportation/ [accessed 2015-04-12] [WebCite Cache ID 6XlpDrNbx]
24. Bayor RH. Roads to racial segregation: Atlanta in the twentieth century. Journal of Urban History 1988 Nov 01;15(1):3-21. [doi: 10.1177/009614428801500101]
25. Doubeni CA, Schootman M, Major JM, Stone Rosalie A Torres, Laiyemo AO, Park Y, et al. Health status, neighborhood socioeconomic context, and premature mortality in the United States: The National Institutes of Health-AARP Diet and Health Study. Am J Public Health 2012 Apr;102(4):680-688 [FREE Full text] [doi: 10.2105/AJPH.2011.300158] [Medline: 21852636]
26. Sampson RJ, Morenoff JD, Gannon-Rowley T. Assessing “neighborhood effects”: Social processes and new directions in research. Annu. Rev. Sociol 2002 Aug;28(1):443-478. [doi: 10.1146/annurev.soc.28.110601.141114]
27. Kramer MR, Hogue CR. Is segregation bad for your health? Epidemiol Rev 2009;31:178-194 [FREE Full text] [doi: 10.1093/epirev/mxp001] [Medline: 19465747]
28. SurveyGizmo. URL: http://www.surveygizmo.com/ [accessed 2015-04-13] [WebCite Cache ID 6Xlpwqfpq]
29. Dasgupta S, Vaughan AS, Kramer MR, Sanchez TH, Sullivan PS. Use of a Google Map tool embedded in an internet survey instrument: Is it a valid and reliable alternative to geocoded address data? JMIR Res Protoc 2014;3(2):e24 [FREE Full text] [doi: 10.2196/resprot.2946] [Medline: 24726954]
30. United States Census Bureau. American fact finder URL: http://factfinder.census.gov/faces/nav/jsf/pages/index.xhtml [accessed 2015-04-13] [WebCite Cache ID 6XlqIax0P]
31. Atlanta Regional Commission. GIS data and maps: Regional transportation URL: http://www.atlantaregional.com/info-center/gis-data-maps/gis-data [accessed 2015-04-13] [WebCite Cache ID 6XlqWnPq9]
32. Summary of travel trends. 2009 National household travel survey: US Department of Transportation, Federal Highway Administration; 2011 URL: http://nhts.ornl.gov/2009/pub/stt.pdf [accessed 2015-04-12] [WebCite Cache ID 6Xlqxl1Hq]
33. Hagman O. Mobilizing meanings of mobility: Car users’ constructions of the goods and bads of car use. Transportation Research Part D: Transport and Environment 2003 Jan;8(1):1-9. [doi: 10.1016/S1361-9209(02)00014-7]

34. Paulley N, Balcombe R, Mackett R, Titheridge H, Preston J, Wardman M, et al. The demand for public transport: The effects of fares, quality of service, income and car ownership. Transport Policy 2006 Jul;13(4):295-306. [doi: 10.1016/j.tranpol.2005.12.004]
35. Hixson BA, Omer SB, del RC, Frew PM. Spatial clustering of HIV prevalence in Atlanta, Georgia and population characteristics associated with case concentrations. J Urban Health 2011 Feb;88(1):129-141 [FREE Full text] [doi: 10.1007/s11524-010-9510-0] [Medline: 21249526]
36. Pucher J, Renne JL. Rural mobility and mode choice: Evidence from the 2001 National Household Travel Survey. Transportation 2005 Mar;32(2):165-186. [doi: 10.1007/s11116-004-5508-3]
37. Dasgupta S, Kramer MR, Rosenberg ES, Sanchez TH, Sullivan PS. Spatial accessibility to HIV providers in Atlanta, Georgia. AIDS Res Hum Retroviruses 2015 May;31(5):473-474. [doi: 10.1089/aid.2014.0365]
38. Liang TS, Erbelding E, Jacob CA, Wicker H, Christmyer C, Brunson S, et al. Rapid HIV testing of clients of a mobile STD/HIV clinic. AIDS Patient Care STDS 2005 Apr;19(4):253-257. [doi: 10.1089/apc.2005.19.253] [Medline: 15857197]
39. Morin SF, Khumalo-Sakutukwa G, Charlebois ED, Routh J, Fritz K, Lane T, et al. Removing barriers to knowing HIV status: Same-day mobile HIV testing in Zimbabwe. J Acquir Immune Defic Syndr 2006 Feb 1;41(2):218-224. [Medline: 16394855]
40. Larson BA, Schnippel K, Ndibongo B, Xulu T, Brennan A, Long L, et al. Rapid point-of-care CD4 testing at mobile HIV testing sites to increase linkage to care: An evaluation of a pilot program in South Africa. J Acquir Immune Defic Syndr 2012 Oct 1;61(2):e13-e17 [FREE Full text] [doi: 10.1097/QAI.0b013e31825eec60] [Medline: 22659650]
41. Sarnquist CC, Soni S, Hwang H, Topol BB, Mutima S, Maldonado YA. Rural HIV-infected women's access to medical care: Ongoing needs in California. AIDS Care 2011 Jul;23(7):792-796. [doi: 10.1080/09540121.2010.516345] [Medline: 21287418]
42. Edgerley LP, El-Sayed YY, Druzin ML, Kiernan M, Daniels KI. Use of a community mobile health van to increase early access to prenatal care. Matern Child Health J 2007 May;11(3):235-239. [doi: 10.1007/s10995-006-0174-z] [Medline: 17243022]
43. Song Z, Hill C, Bennet J, Vavasis A, Oriol NE. Mobile clinic in Massachusetts associated with cost savings from lowering blood pressure and emergency department use. Health Aff (Millwood) 2013 Jan;32(1):36-44 [FREE Full text] [doi: 10.1377/hlthaff.2011.1392] [Medline: 23297269]
44. Sampson R. Rail Magazine. 2011. Carpe Diem: Georgia prepares to seize its mobility future URL: http://web1.ctaa.org/webmodules/webarticles/articlefiles/Fall-Winter11_Georgia.pdf [accessed 2015-04-13] [WebCite Cache ID 6Xls1si35]

Abbreviations

HIV: human immunodeficiency virus
IQR: interquartile ranges
MARTA: Metropolitan Atlanta Rapid Transit Authority
MSM: men who have sex with men
SES: socioeconomic status

Edited by G Eysenbach; submitted 14.04.15; peer-reviewed by C George, S Stahlman, B Linas; comments to author 26.06.15; revised version received 19.07.15; accepted 27.07.15; published 24.08.15.

Please cite as:
Dasgupta S, Kramer MR, Rosenberg ES, Sanchez TH, Reed L, Sullivan PS
The Effect of Commuting Patterns on HIV Care Attendance Among Men Who Have Sex With Men (MSM) in Atlanta, Georgia
JMIR Public Health Surveill 2015;1(2):e10
URL: http://publichealth.jmir.org/2015/2/e10/
doi:10.2196/publichealth.4525
PMID:27227128

Original Paper

An Integrated Service Delivery Model to Identify Persons Living with HIV and to Provide Linkage to HIV Treatment and Care in Prioritized Neighborhoods: A Geotargeted, Program Outcome Study

Paula M Frew[1,2], MA, MPH, PhD; Matthew Archibald[3], PhD; Jay Schamel[1], BS; Diane Saint-Victor[1], MPH; Elizabeth Fox[1]; Neena Smith-Bankhead[4], MS; Dazon Dixon Diallo[5], MPH; Marcia M Holstad[6], PhD, FNP-BC, FAAN, FAANP; Carlos del Rio[1,2], MD

[1]Division of Infectious Diseases, Department of Medicine, Emory University School of Medicine, Atlanta, GA, United States

[2]Hubert Department of Global Health, Rollins School of Public Health, Emory University, Atlanta, GA, United States

[3]Department of Sociology, Colby College, Waterville, ME, United States

[4]AID Atlanta, Atlanta, GA, United States

[5]SisterLove, Inc., Atlanta, GA, United States

[6]Emory University School of Nursing, Atlanta, GA, United States

Corresponding Author:
Paula M Frew, MA, MPH, PhD
Division of Infectious Diseases
Department of Medicine
Emory University School of Medicine
1760 Haygood Drive, Suite W327
Atlanta, GA, 30322
United States
Phone: 1 404 712 8546
Fax: 1 404 370 6142
Email: pfrew@emory.edu

Abstract

Background: Recent studies have demonstrated that high human immunodeficiency virus (HIV) prevalence (2.1%) rates exist in “high-risk areas” of US cities that are comparable to rates in developing nations. Community-based interventions (CBIs) have demonstrated potential for improving HIV testing in these areas, thereby facilitating early entry and engagement in the HIV continuum of care. By encouraging neighborhood-based community participation through an organized community coalition, Project LINK sought to demonstrate the potential of the CBI concept to improve widespread HIV testing and referral in an area characterized by high poverty and HIV prevalence with few existing HIV-related services.

Objective: This study examines the influence of Project LINK to improve linkage-to-care and HIV engagement among residents of its target neighborhoods.

Methods: Using a venue-based sampling strategy, survey participants were selected from among all adult participants aged 18 years or more at Project LINK community events (n=547). We explored multilevel factors influencing continuum-of-care outcomes (linkage to HIV testing and CBI network referral) through combined geospatial-survey analyses utilizing hierarchical linear model methodologies and random-intercept models that adjusted for baseline effect differences among zip codes. The study specifically examined participant CBI utilization and engagement in relation to individual and psychosocial factors, as well as neighborhood characteristics including the availability of HIV testing services, and the extent of local prevention, education, and clinical support services.

Results: Study participants indicated strong mean intention to test for HIV using CBI agencies (mean 8.66 on 10-point scale [SD 2.51]) and to facilitate referrals to the program (mean 8.81 on 10-point scale [SD 1.86]). Individual-level effects were consistent across simple multiple regression and random-effects models, as well as multilevel models. Participants with lower income expressed greater intentions to obtain HIV tests through LINK (P<.01 across models). HIV testing and CBI referral intention were associated with neighborhood-level factors, including reduced availability of support services (testing P<.001),

greater proportion of black/African Americans (testing and referral $P<.001$), and reduced socioeconomic capital (testing $P=.017$ and referral $P<.001$). Across models, participants expressing positive attitudes toward the CBI exhibited greater likelihood of engaging in routine HIV testing ($P<.01$) and referring others to HIV care ($P<.01$). Transgender individuals indicated greater intent to refer others to the CBI ($P<.05$). These outcomes were broadly influenced by distal community-level factors including availability of neighborhood HIV support organizations, population composition socioeconomic status, and high HIV prevalence.

Conclusions: Project LINK demonstrated its potential as a geotargeted CBI by evidencing greater individual intention to engage in HIV testing, care, and personal referrals to its coalition partner organizations. This study highlights important socioecological effects of US-based CBIs to improve HIV testing and initiate acceptable mechanisms for prompt referral to care among a vulnerable population.

(JMIR Public Health Surveill 2015;1(2):e16) doi:10.2196/publichealth.4675

KEYWORDS

human immunodeficiency virus; human immunodeficiency virus prevention; human immunodeficiency virus testing; racial/ethnic minorities; community-based organizations; High-Impact Prevention

Introduction

Background

Community-based interventions (CBIs) are a feasible, sustainable approach to increase widespread human immunodeficiency virus (HIV) testing and improve entry and engagement in the HIV continuum of care [1,2]. Ideally, engagement in care is a seamless, coordinated process commencing with individual testing, diagnosis, and treatment initiation. Yet those at highest risk of HIV infection are the most challenging to engage and susceptible to delays across the care continuum. HIV testing delay is frequent in US men-who-have-sex-with-men (MSM) populations, with an estimated 19-26% of MSM unaware of their status [3]. Delayed testing is associated with a lack of awareness or denial of perceived risk for infection, age, and race/ethnicity [3,4]. Racial and ethnic minorities are at increased risk of delayed referral to HIV care and treatment following diagnosis [5,6]. Rates of delayed testing rates remain high; in 2013, 23.6% of newly diagnosed HIV patients in the United States were classified as Stage 3 (acquired immune deficiency syndrome, AIDS) at diagnosis [7]. Treatment delay is more common among black/African Americans, immigrants, and uninsured individuals [8]. Between 20% and 40% do not link to HIV care within a year of diagnosis, a delay which is associated with higher rates of virologic failure, increased morbidity and mortality, and immune system damage resulting from delayed receipt of antiretrovirals [9].

Recent studies have demonstrated that high HIV prevalence (2.1%), comparable with HIV rates in developing nations, are present in “high-risk areas” of US cities, particularly in neighborhoods characterized by high poverty and HIV prevalence [10]. These “hot spot” areas are experiencing local, yet generalized, HIV microepidemics. Notably, many of these areas are located within 12 major metropolitan areas that account for approximately 44% of all estimated AIDS cases, signifying the challenges facing continuum of care access and delivery [11]. Thus, it is important to look at health care service delivery in these areas and examine the extent to which these options are culturally compatible and socially sensitive to the needs of those who could most benefit from geographically targeted HIV prevention and care. Previous studies have identified the importance of recognizing the spatial distribution of HIV burden [11,12], HIV service provision and continuum of care objectives [13-15], and also of the spatial and ecosocial dimensions of the development and delivery of CBIs targeting HIV transmission [1,16,17].

Project LINK

Project LINK was an initiative supported by the Atlanta AIDS Partnership Fund and the Community Foundation of Greater Atlanta to increase HIV testing in an area of Atlanta, Georgia, characterized by high poverty and HIV prevalence. LINK’s goals were to identify residents living with HIV and directly connect those living with HIV to appropriate medical care and treatment programs. In addition, LINK created a model for building lasting partnerships between community and HIV/AIDS outreach agencies. The project was initiated as a result of numerous meetings with community partners and local residents concerned about the high HIV prevalence rate in their neighborhood. Collectively, all parties reviewed HIV/AIDS data, held discussions on community needs and assets, and worked to identify specific strengths and potential contributions of the selected agency partners to the delivery of HIV prevention and care in the selected neighborhoods.

Development of the intervention thus occurred through a process of community-based participatory research [18]. Community members and leaders were invited to attend a series of meetings with the funder, technical advisors, and evaluative team to discuss factors that may be influencing high HIV prevalence rates in the target neighborhoods for the intervention, a process that has proven to be effective in eliciting critical intervention points [19]. These conversations led to an inventory of structural, social, and individual-level factors that aligned well with the socioecological model [20]. Thus, the intervention was informed by this theoretical framework based on community consensus and resulting activities that focused on addressing such factors across levels. As a result of these planning activities, 5 community partner agencies were selected to collaborate with the selected community to increase the capacity of local residents and agencies to conduct HIV prevention and linkage-to-care activities: the Center for Black Women’s Wellness, Inc (CBWW), the Atlanta Harm Reduction Coalition (AHRC), AID Atlanta, Positive Impact, and SisterLove, Inc.

High-Impact Prevention [21], an intervention approach adopted by the Centers for Disease Control and Prevention (CDC) in 2013 to improve the effectiveness of HIV prevention efforts, maximizes the relevance of proposed interventions through consideration of the cost effectiveness and feasibility of full-scale implementation, as well as assessment of coverage, interactions, and combinations. Project LINK utilized this same strategy in bringing together various types of service providers while supporting partnership development and linkage systems to ensure that those in need of HIV services had access to a comprehensive range of resources appropriate for their needs, accessible to those at greatest risk. Each of the participating agencies had a long history of providing culturally competent programs and integrative service delivery for HIV/AIDS prevention, testing, counseling, and linkage to HIV case managers and mental health services. Each agency had a specific role related to HIV testing and subsequent coordinated referral and linkage to in-house clinical care. In addition, supportive programs such as the evidence-based “HealthyLove” HIV prevention party (SisterLove) [22,23]; needle exchange (AHRC) [24,25]; women’s support groups (CBWW, SisterLove, and AID Atlanta) [26]; and mental health counseling and domestic violence prevention and response training (Positive Impact) [27] were components of the agencies’ service delivery package. These were offered to residents at neighborhood schools, churches, housing complexes, and community meeting locations, in addition to street outreach conducted by the agencies’ mobile units.

The community partners developed Project LINK to normalize routine HIV testing and to provide coordinated linkage to HIV care among the participating partner entities serving residents in the target neighborhoods. LINK focused on engaging residents living in an area comprising 2 target zip codes within a generalized HIV high-prevalence cluster (≥1.0% HIV case prevalence) and few established HIV continuum-of-care resources [12]. For this study, we also identified a secondary catchment area of 5 adjacent zip codes within the cluster, directly abutting the 2 target zip codes. The initiative sought to reach about 5000 residents over a 1-year period.

Methods

Study Design and Sample

The purpose of this study is to assess the influence of the Project LINK CBI to improve linkage-to-care and HIV engagement in a selected area of Atlanta that demonstrated a high HIV prevalence burden and limited health care services in its census tracts. Individuals enrolled in this study were selected from survey sessions that were randomly scheduled over a 10-month demonstration period. This resulted in questionnaire collection at 31 unique LINK-organized community activities and programs during this period. Outcomes related to intention to utilize CBI HIV testing resources and to refer other CBIs were included in surveys administered at 23 of these postimplementation activities within the process evaluation phase of the CBI. Eligibility criteria specified inclusion of men and women aged 18 and over who had the ability to read and write English. Respondents selected a gift card, t-shirt, transit card, or some other health-promotion item for their participation. The Emory University Institutional Review Board approved the protocol (00005278) prior to study inception.

Study Measures

The questionnaire included items on HIV testing and referral to the CBI, in addition to sociodemographics, perceptions, and attitudinal factors. Our primary outcome of interest was participants’ willingness to engage in routine HIV testing through the CBI, measured by the item, "On a scale from 0 (definitely not) to 10 (definitely so), rank your likelihood of getting your next HIV test with a LINK agency in the next 6 months." To assess the potential of the CBI to motivate participants to refer others to its services, we included the item, "On a scale from 0 (definitely not) to 10 (definitely so), rank your likelihood of getting others involved in Project LINK."

Individual-Level Factors

We estimated the impact of multiple, nested influential factors on the CBI’s key outcomes of interest, HIV testing, and LINK service referral. Demographic factors from survey responses included race, income, gender, and age. For the purposes of this study, race was coded as "white" and "nonwhite." Household income responses included 6 categories in US $20,000 increments from “less than US $20K per year” to “over US $100K per year,” and were treated as a continuous variable for our analyses. Gender was included in our models as a three-level categorical variable with levels "male," "female," and "transgender." Age was measured in years, and was incorporated into our models as a continuous covariate.

Exploratory Factor Analysis

Psychometric items were drawn from previous HIV behavioral research demonstrating excellent internal consistency and validity of items [20,28]. The questionnaire included measures assessing attitudes about the LINK initiative, HIV/AIDS, social norms, and community involvement [1,29,30]. Subscale development utilized principal component analyses with varimax rotation, followed by assessment of components’ internal consistency. Because some item responses were missing (3.6%), we conducted Little’s Missing Completely at Random (MCAR) test to analyze the overall pattern on surveys and subsequently performed mean imputation.

Neighborhood-Level Factors

To assess the CBI’s success in promoting linkage to HIV support in underserved communities, we explored the relationship between LINK service utilization and the availability of HIV services, support, and educational organizations. Previous research provided a comprehensive catalog of HIV service offerings in both the 2 CBI target area zip codes and the surrounding neighborhoods [12]. Using their reported residential zip code, each participant was connected to the total number of discrete, permanent HIV-related services available in their community [13]. Discrete services included HIV case management, HIV medical treatment/services, HIV testing and counseling, community education and outreach, mentoring and support services, etc. and separate services from the same provider were counted separately. Only services provided prior to the initiation of Project LINK were included in this count.

HIV prevalence was estimated for each zip code using census-tract-level HIV diagnosis counts from 2005 to 2007. These census-tract-level HIV counts were aggregated to zip-code-level counts using Esri ArcGIS version 10.2 [31]. Counts from census tracts overlapping more than 1 zip code were split by area. HIV prevalence was computed by dividing the aggregate HIV diagnosis count by the zip code population, as measured in the US Census 2000 [32].

Other neighborhood-level factors were included to reflect the socioeconomic composition of the community. These variables included the proportion of black/African American residents, the proportion of residents aged 25 years or more, the proportion of male residents over 18 who have graduated high school, median income, male employment rate, and the proportion of vacant households. These community characteristics were obtained at the zip code level from the US Census Bureau's Census 2000 [32].

Statistical Analyses

We first computed descriptive statistics for characteristics of our sample of CBI participants and for questions eliciting participant impressions of the CBI. We then computed descriptive statistics for our 2 outcome measures, willingness to engage in routine HIV testing through the CBI, and intention to refer others to the CBI. To compare these outcomes between participants living in the 2 primary target zip codes, those living in the 5 secondary catchment zip codes, and those living outside the target areas, we utilized analysis of variance (ANOVA) post hoc pairwise analysis with Tamhane adjustment.

Next, we employed random-intercept linear mixed models to examine the effect of individual- and neighborhood-level covariates on CBI utilization and referral outcomes. Baseline differences among participants from different zip codes were adjusted for through the incorporation of random intercepts for each zip code. To focus analysis on effects relating to individuals within the CBI area, only participants from the 2 target and 5 secondary catchment zip codes were included in the multilevel analysis. Because 7 zip codes did not admit multiple neighborhood effects in a single model, separate models were fit for each neighborhood-level covariate, each regressing a single neighborhood-level covariate and all individual-level covariates on a CBI outcome. To assess the stability of individual-level effects, multiple linear and random-intercept (by zip code) models were also fit using only the individual and psychosocial variables, excluding neighborhood-level variables. Random-intercept models used the *xtreg* procedure with maximum likelihood estimation in Stata version 13 [33]. Participants with missing outcome responses were excluded by listwise deletion. Variance inflation factors were used to assess all models for multicollinearity; no issues were discovered. For all hypothesis tests, results were considered statistically significant if $P<0.05$.

Results

Sample Characteristics

Of the 597 respondents selected at the 23 postimplementation activities, 414 (69%) lived within the 2 primary LINK target zip codes, 37 (6.2%) within the 5 secondary catchment zip codes, 101 (17%) lived outside the targeted area, and 45 (7.5%) did not list a home zip code. Table 1 describes the sociodemographic characteristics of the sampled participants, together with the characteristics of the participants living within the 2 target zip codes and the 5 secondary catchment zip codes (Table 1). The CBI participants included a majority of black/African American (88.8%, n=530) participants in the age range of 40-59 years (63.7%, n=380; Table 1). Respondents were evenly split between male and female participants (47.6%, n=284 versus 45.2%, n=270). In addition, the sample included 27 transgender persons (the majority male-to-female). Most respondents obtained high-school diplomas or general educational developments (56.8%, n=339), yet many were also unemployed (54.6%, n=326) and had annual household income less than US $20,000 per year (78.2%, n=467).

Table 1. Participant sociodemographic characteristics.[a]

		All respondents	Target area and secondary catchment
		Frequency (%)	Frequency (%)
Age [b,c]			
	18-29 years	78 (13.1)	56 (12.4)
	30-39 years	85 (14.2)	58 (12.9)
	40-49 years	213 (35.7)	175 (38.8)
	50-59 years	167 (28.0)	133 (29.5)
	≥60 years	35 (5.9)	23 (5.1)
	Missing	*19 (3.2)*	*6 (1.3)*
Gender			
	Male	284 (47.6)	218 (48.3)
	Female	270 (45.2)	207 (45.9)
	Transgender: M to F	22 (3.7)	15 (3.3)
	Transgender: F to M	5 (0.8)	3 (0.7)
	Missing	*16 (2.7)*	*8 (1.8)*
Race			
	White	29 (4.9)	12 (2.7)
	Nonwhite	531 (88.9)	415 (92.0)
	Missing	*37 (6.2)*	*43 (9.5)*
Ethnicity			
	Asian/Asian American/Pacific Islander	11 (1.8)	5 (1.1)
	Hispanic/Latino/Chicano	1 (0.2)	0 (0.0)
	African American/black	530 (88.8)	421 (93.3)
	Caucasian/white	20 (3.4)	6 (1.3)
	American Indian/Alaska Native	6 (1.0)	3 (0.7)
	Multiracial/Multicultural	10 (1.7)	7 (1.6)
	Missing	*19 (3.2)*	*9 (2.0)*
Sexual orientation			
	Heterosexual	512 (85.8)	400 (88.7)
	LGBTQQ[d]	69 (11.6)	42 (9.3)
	Missing	*16 (2.7)*	*9 (2.0)*
Relationship status			
	Single	387 (64.8)	299 (66.3)
	Married/domestic partner	105 (17.6)	67 (14.9)
	Divorced/separated	69 (11.6)	60 (13.3)
	Widowed	26 (4.4)	21 (4.7)
	Missing	*10 (1.7)*	*4 (0.9)*
Educational attainment			
	K-8 grade	71 (11.9)	52 (11.5)
	High-school graduate/general educational development	339 (56.8)	282 (62.5)
	Technical/Vocational or Associates	100 (16.8)	81 (18.0)

		All respondents	Target area and secondary catchment
	Bachelor degree	39 (6.5)	16 (3.5)
	Master's degree	14 (2.3)	3 (0.7)
	Doctorate	9 (1.5)	1 (0.2)
	Missing	*25 (4.2)*	*16 (3.5)*
Household income			
	Less than US $20,000	467 (78.2)	395 (87.6)
	US $20,001-US $40,000	43 (7.2)	25 (5.5)
	US $40,001-US $60,000	19 (3.2)	5 (1.1)
	US $60,001-US $80,000	16 (2.7)	4 (0.9)
	US $80,000-US $100,000	9 (1.5)	1 (0.2)
	More than US $100,000	17 (2.8)	4 (0.9)
	Missing	*26 (4.4)*	*17 (3.8)*
Employment status			
	Employed full time	91 (15.2)	39 (8.6)
	Employed part-time	95 (15.9)	80 (17.7)
	Unemployed	326 (54.6)	279 (61.9)
	Other[e]	73 (12.2)	49 (10.9)
	Missing	*12 (2.0)*	*4 (0.9)*
Distance traveled to CBI activity			
	<5 miles	459 (76.9)	388 (86.0)
	6-9 miles	57 (9.5)	33 (7.3)
	10-20 miles	37 (6.2)	10 (2.2)
	>20 miles	31 (5.2)	14 (3.1)
	Missing	*13 (2.2)*	*6 (1.3)*

[a]The italics are used to emphasize the percentage of nonresponders for each item

[b]Mean and SD for all respondents: 44.6 and 11.4, respectively.

[c]Mean and SD for respondents in target area and secondary catchment: 44.9 and 11.0, respectively.

[d]Lesbian, gay, bisexual, transgender, queer, and questioning

[e]Retired, student, self-employed, disability, and illicit

Individuals in our sample provided insight on their motivations to attend the LINK activities (Table 2). Most common reasons provided were to obtain medical and scientific information (36.5%, n=218) and meeting others who share similar HIV/AIDS concerns in the community (25.3%, n=151). A majority of respondents also expressed strong approval of LINK activities (64.0%, n=382) and 57.0% rated Project LINK as excellent/outstanding (n=340).

Table 2. Participant CBI impressions.[a]

	All respondents (n=597)	Target area and secondary catchment (n=451)
	Frequency (%)	Frequency (%)
Motivation to attend CBI activity		
Get the latest scientific/medical information	218 (36.5)	168 (37.3)
Meet others who share my concerns	151 (25.3)	119 (26.4)
Sense of obligation to my community	88 (14.7)	58 (12.9)
Learn about volunteer opportunities	58 (9.7)	46 (10.2)
Other[b]	52 (8.7)	42 (9.3)
Missing	*30 (5.0)*	*18 (4.0)*
Rating of this event/activity		
Excellent/outstanding	382 (64.0)	302 (67.0)
Good/very good	179 (30.0)	131 (29.0)
Fair/poor	8 (1.3)	6 (1.3)
No opinion	10 (1.7)	4 (0.9)
Missing	*18 (3.0)*	*8 (1.8)*
Overall impression of Project LINK		
Excellent/outstanding	340 (57.0)	268 (59.4)
Good/very good	226 (37.9)	167 (37.0)
Fair/poor	6 (1.0)	5 (1.1)
No opinion	15 (2.5)	8 (1.8)
Missing	*10 (1.7)*	*3 (0.7)*

[a]The italics are used to emphasize the percentage of nonresponders for each item.

[b]Multiple responses given included write-in comments such as accompanying friend/relative, treatment/testing, compensation, life change, and community service.

The 2 primary CBI target zip codes and 5 secondary catchment zip codes differed in the availability of HIV continuum-of-care services. Within 1 primary target zip code, there were no HIV services available to residents. The other primary zip code had 10 HIV services located within the area (eg, testing, support). The number of available HIV services identified within the 5 secondary target zip codes ranged from only 3 or 4 in 3 of the zip code areas to 49 in 1 zip code area.

Psychosocial Factors

Psychosocial subscale items and results of the exploratory factor analysis are detailed in Table 3. Chosen subscales include *LINK Attitudes* about the risks of HIV and benefits of LINK involvement, degree of psychological *LINK Engagement*, *Negative Participatory Norms* associated with HIV testing and the CBI, *Perceived LINK Social Support*, and *Identification with LINK Brand* (Table 3). The scales exhibited excellent psychometric properties including strong internal consistencies across domains (alpha=.733-.940). Responses to the items were rated on a 5-point scale and subscale scores were summed; higher scores indicated higher levels of the attribute.

Table 3. Factor subscales for psychosocial domains.

Factor	Factor characteristics and questions	Loading
LINK attitudes[a]		
	I benefit from Project LINK services.	.77
	I like getting involved with Project LINK.	.76
	My community will really benefit from Project LINK.	.73
	My involvement will improve my community's trust in Project LINK.	.73
	My involvement in Project LINK will improve my health.	.73
	My participation in Project LINK would be very good.	.72
	I would participate in Project LINK activities because it would help to prevent AIDS.	.72
	I feel that my involvement in Project LINK is making an important difference.	.70
	HIV testing is a benefit of getting involved.	.68
	HIV is a serious concern in my immediate community.	.68
LINK engagement[b]		
	Getting involved in the Project LINK effort is liberating.	.77
	Project LINK is a social justice effort.	.74
	Project LINK will reduce health disparities.	.70
	I feel a sense of purpose in this cause.	.69
	It is fun to be involved with the Project LINK.	.68
	I feel a sense of belonging through my participation in this effort.	.67
	My involvement is helping to protect the rights of others.	.67
	I am advancing the public's health and well-being through my support of this cause.	.65
	I am as source of HIV/AIDS information in my community.	.62
	Being involved with the Project LINK helps me to feel empowered.	.59
	I experience a sense of community in this cause.	.59
	I would be very concerned about the outcome of any effort of which I am affiliated.	.42
	It is extremely important to make the right choice in selecting a volunteer organization.	.40
	The Project LINK effort is very different from others.	.40
Negative participatory norms[c]		
	I think my friends would negatively judge me if I sought HIV testing.	.84
	I tend to be worried about what people think of me, even if I do not know them.	.79
	Participating in Project LINK seems risky.	.75
	I think some of my family members would be upset if I participated in Project LINK.	.72
	People negatively judge those who participate in Project LINK.	.70
	Even if I wanted to participate in Project LINK, I just do not have the time.	.69
	I generally do what my family expects of me.	.58
	If people heard of my participation with the Project LINK, they would form an opinion of me.	.57
	In general, I am among the last of my circle of friends to try new things.	.55
LINK social support[d]		
	If I decided to participate in Project LINK, I probably would tell my partner.	.61
	I would do something even if members of my social group disagreed with my actions.	.56
	I think my work colleagues would approve of my involvement.	.54

Factor	Factor characteristics and questions	Loading
	Most people important to me think my involvement in Project LINK is good.	.50
	I think my doctor would approve of my involvement in Project LINK.	.49
	My immediate family is supportive of my involvement in Project LINK.	.48
	If my pastor supported Project LINK, I would be inclined to get involved.	.42
LINK brand perception[e]		
	Prior to joining any organization, I prefer to consult a friend who has experience with that group.	.64
	Hearing that somebody else is involved with the Project LINK tells me a lot about that person.	.58
	When it comes to deciding whether to join a new organization, I rely on experienced friends or family members for advice.	.56
	Being active with the Project LINK would help me to express who I am.	.47
	You can tell a lot about a person by their community affiliations.	.37

[a]Alpha=0.940; 10 items

[b]Alpha=0.935; 14 items

[c]Alpha=0.880; 9 items

[d]Alpha=0.830; 7 items

[e]Alpha=0.733; 5 items

Linkage to HIV Testing

Most participants felt comfortable obtaining an HIV test with designated LINK providers, as indicated by high mean intention to test for HIV using a LINK agency (mean 8.66 on 10-point scale [SD 2.51]). One-way ANOVA by residence within CBI target areas (primary, secondary, outside of target area) found statistically significant differences among persons living in the LINK target area, those living adjacent to the primary intervention zone, and those coming from outside the designated zip codes with desire to use CBI HIV testing resources ($F_{2,447}$=11.6, P<.001). Tamhane post hoc analyses indicated that respondents living in the 2 CBI target zip codes expressed greater intention to engage in routine HIV testing through the CBI compared with those living outside the target and secondary catchment zip codes (difference=1.6, P=.004).

The results of the multiple regression and random-intercept models with individual-level covariates are presented in Table 4 and model parameters in Table 5. Figure 1 shows the individual predictors of HIV testing. The multiple regression model incorporates all individual and psychosocial independent variables (race, income, gender, age, and the 5 psychosocial scales), but no neighborhood-level factors. The individual-level random-intercept model adds a random intercept for zip code to the previous model. Results from the multilevel models containing all individual-level covariates and a single neighborhood-level covariate are detailed in Table 6. Figure 2 shows the adjusted neighborhoods predictors of HIV testing. Full results for individual and psychosocial effects in the multilevel models are given in Multimedia Appendix 1. The coefficient estimates for individual and psychosocial covariates are very similar across multivariable models (Tables 4 and 6). Among the demographic covariates, lower income is associated with greater willingness to test in all models (P<.01 across models). Older individuals have increased testing intention in the multiple linear model (P=.02). Among the psychosocial factors, favorable “LINK attitudes” (P<.01 across models), “LINK engagement” (P<.05 across models), and “Identification with LINK brand” (P<.05 across model) were all associated with increased desire to obtain routine HIV testing through LINK.

Within the target area and secondary catchment, participants living in zip codes with fewer available HIV services (P<.001), lower HIV prevalence (P=.01), greater proportion of black/African Americans (P<.001), smaller proportion of those 24 years or older living in the community (P<.001), and reduced median household income (P=.02) were associated with increased desire to obtain HIV testing through LINK, after adjustment for individual factors (Table 5).

Table 4. Individual-level predictor models for HIV testing.

	Multiple linear model			Random intercept model		
Predictor	Coefficient (95% CI)	Standardized coefficient (95% CI)	*P* value	Coefficient (95% CI)	Standardized coefficient (95% CI)	*P* value
Intercept	8.41 (5.86 to 10.97)	8.92 (8.70 to 9.13)	<.001	8.12 (5.65 to 10.75)	8.72 (8.05 to 9.38)	<.001
Individual						
Race (white ref)	0.07 (-1.13 to 1.26)	0.01 (-0.20 to 0.22)	.91	0.14 (-1.02 to 1.31)	0.02 (-0.18 to 0.23)	.81
Income[a]	-0.59 (-0.97 to -0.21)	-0.40 (-0.67 to 0.14)	.003	-0.60 (-0.97 to -0.23)	-0.41 (-0.66 to 0.16)	.002
Gender (Male ref)						
Female	0.20 (-0.25 to 0.64)	0.10 (-0.12 to 0.32)	.38	0.22 (-0.21 to 0.66)	0.11 (-0.11 to 0.33)	.32
Transgender	0.83 (-0.29 to 1.94)	0.17 (-0.06 to 0.39)	.15	0.71 (-0.38 to 1.80)	0.14 (-0.08 to 0.36)	.20
Age (years)	0.02 (0.01 to 0.40)	0.23 (0.01 to 0.44)	.04	0.02 (-0.02 to 0.37)	0.19 (-0.02 to 0.41)	.08
Psychosocial						
LINK attitudes	0.38 (0.16 to 0.60)	0.36 (0.15 to 0.58)	<.001	0.35 (0.13 to 0.57)	0.34 (0.13 to 0.55)	.002
LINK engagement	0.31 (0.09 to 0.53)	0.31 (0.09 to 0.53)	.006	0.30 (0.09 to 0.52)	0.30 (0.08 to 0.52)	.006
Negative participatory norms	-0.00 (-0.23 to 0.22)	-0.00 (-0.23 to 0.22)	.98	-0.03 (-0.25 to 0.19)	-0.03 (-0.25 to 0.19)	.78
LINK social support	0.09 (-0.14 to 0.32)	0.09 (-0.14 to 0.32)	.44	0.09 (-0.14 to 0.31)	0.08 (-0.14 to 0.31)	.45
LINK Brand Perc.	-0.02 (-0.24 to 0.21)	-0.02 (-0.23 to 0.20)	.89	0.00 (-0.21 to 0.22)	0.00 (-0.21 to 0.22)	.97

[a]Income recorded in US $20K categories from 1 (<US $20,000) to 6 (>US $100,000).

Table 5. Model parameters for individual-level predictor models for HIV testing.

Model	N	Listwise deleted	σ_u (95% CI)	σ_e (95% CI)	Rho (95% CI)	LR test of σ_u=0 (*P*)	AIC	BIC
Multiple linear model	421	63	—	—	—	—	1543	1586
Random intercept model	421	63	0.54 (0.09 to 3.14)	2.00 (1.86 to 2.16)	0.07 (0.00 to 0.59)	.05	1545	1595

Table 6. Multilevel predictor models for HIV Testing.[a]

	Neighborhood variable			Model parameters		
Adjusted Neighborhood-Level Predictor[b]	Coefficient (95% CI)	Standard. Coefficient (95% CI)	*P*	Rho (95%CI)	LR test of σ_u=0 (*P*)	AIC
Number of HIV support services	-0.05 (-0.08 to -0.03)	-0.39 (-0.59 to -0.19)	<.001	0	>0.99	1535
HIV prevalence	-0.61 (-0.11 to -0.01)	-0.32 (-0.58 to -0.06)	.01	0.03 (0.00 to 0.16)	.008	1541
Black/African American population (%)	0.02 (0.01 to 0.04)	0.36 (0.14 to 0.57)	<.001	0	1.00	1539
Age ≥ 25 years (%)	-0.08 (-0.13 to -0.03)	-0.39 (-0.62 to -0.15)	<.001	0	1.00	1539
Male high-school graduation rate (%)	-0.08 (-0.17 to 0.00)	-0.26 (-0.53 to 0.00)	.05	0.02 (0.00 to 0.16)	.07	1543
Male employment (%)	-0.03 (-0.06 to 0.01)	-0.25 (-0.55 to 0.06)	.11	0.01	.14	1543
Median household income (US $K)	-0.07 (-0.13 to -0.01)	-0.34 (-0.61 to 0.06)	.02	0.01	.16	1540
Vacant homes (%)	0.04 (-0.06 to 0.14)	0.16 (-0.21 to 03.53)	.41	0.01	.12	1545

[a]Each multilevel model includes a single neighborhood-level variable together with all individual and psychosocial variables (N=421; 63 listwise deleted).

[b]Adjusted for race, income, gender, age, and the 5 psychosocial variables

Figure 1. Individual predictors of HIV Testing.

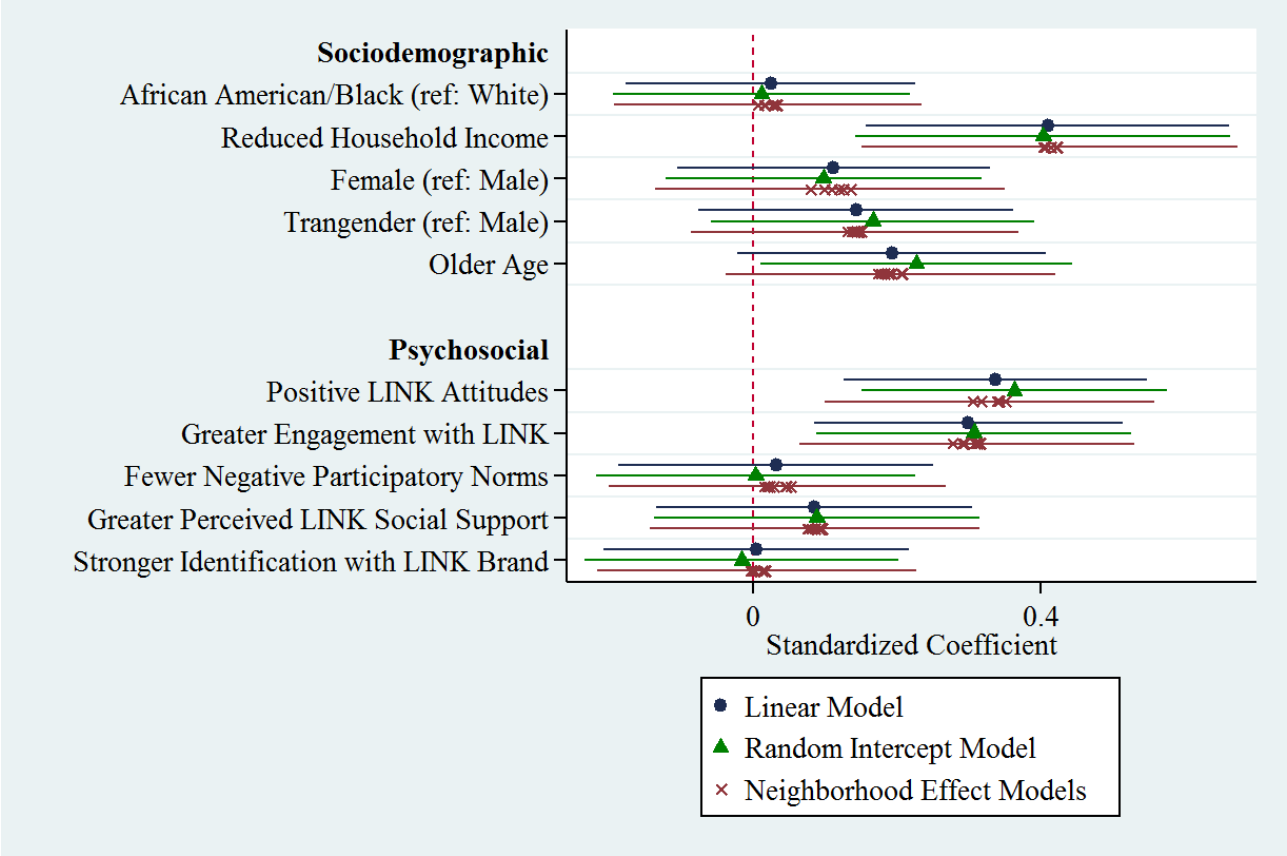
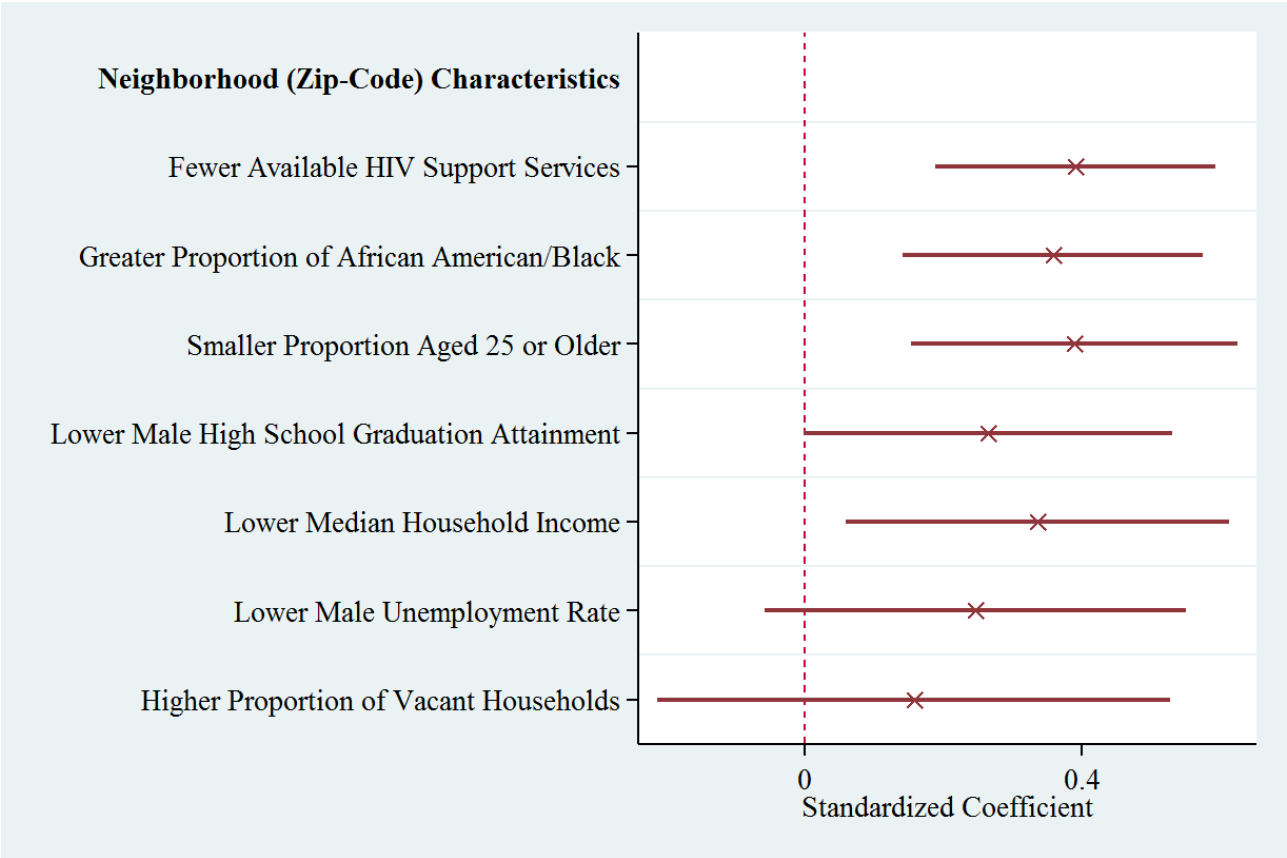


Figure 2. Adjusted neighborhoods predictors of HIV Testing.



Participant Network Referral

Along with HIV testing intentions, participants also expressed a strong desire to refer other persons to LINK (mean 8.81 on 10-point scale [SD 1.86]). One-way ANOVA found significant differences in willingness to refer others among participants living in different areas ($F_{2,397}$=5.812, P=.003). Tamhane post hoc analysis indicated that participants residing in the 2 CBI target zipcodes expressed greater intention to refer others to LINK than those residing outside the target and secondary catchment zip codes (difference=0.8, P=.01).

The results of individual-level multivariable and random-intercepts models for CBI referral are detailed in Table 7 and model parameters in Table 8; multilevel neighborhood-factor models are provided in Table 9. Figures 3 and 4 show individual predictors of CBI referral and adjusted neighborhoods predictors of CBI referral, respectively. All models incorporate all individual and psychosocial covariates (race, income, gender, age, and the 5 psychosocial scales). The random-intercepts model further includes a random intercept for zip code, and the multilevel neighborhood-factor models include random intercepts as well as 1 neighborhood factor in each model. Full results for individual and psychosocial covariates in the multilevel models are given in Multimedia Appendix 1. The likelihood ratio test for null variance of the random intercept was statistically significant for referral to LINK services (P<.01 across models; Table 7). Individual-level coefficients are similar among these models (Tables 7 and 9). The results demonstrate that transgender individuals were more willing than men to refer members of their social network to LINK services (P<.05 across models). Participants living in zip codes with greater proportion of black/African Americans living in the CBI area (P<.001), greater proportion of vacant homes (P=.002), and reduced median household income (P<.001) were associated with increased desire to initiate referrals to LINK.

Table 7. Individual-level predictor models for HIV service referral.

	Multiple linear model			Random intercept model		
Predictor	Coefficient (95% CI)	Standardized Coefficient (95% CI)	*P*	Coefficient (95% CI)	Standardized Coefficient (95% CI)	*P*
Intercept	9.07 (6.79 to 11.35)	8.90 (8.71 to 9.10)	<.001	8.94 (6.72 to 11.17)	8.89 (8.49 to 9.30)	<.001
Individual						
Race (white ref)	-0.32 (-0.41 to 0.76)	-0.06 (-0.24 to 0.13)	.56	-0.21 (-1.25 to 0.84)	-0.04 (-0.22 to 0.14)	.70
Income[a]	-0.10 (-0.36 to 2.39)	-0.07 (-0.25 to 0.12)	.48	-0.10 (-0.35 to 1.67)	-0.06 (-0.24 to 0.11)	.49
Gender (male ref)						
Female	0.28 (-0.12 to 0.67)	0.14 (-0.06 to 0.34)	.17	0.35 (-0.03 to 0.74)	0.18 (-0.02 to 0.37)	.07
Transgender	1.35 (0.31 to 2.39)	0.27 (0.06 to 0.48)	.01	1.28 (0.28 to 2.28)	0.26 (0.06 to 0.46)	.01
Age (years)	0.08 (-0.10 to 0.26)	0.09 (-0.11 to 0.29)	.40	0.05 (-0.13 to 0.22)	0.05 (-0.14 to 0.24)	.61
Psychosocial						
LINK attitudes	0.32 (0.11 to 0.53)	0.31 (0.11 to 0.51)	.003	0.30 (0.10 to 0.51)	0.29 (0.10 to 0.49)	.004
LINK engagement	0.36 (0.16 to 0.56)	0.36 (0.16 to 0.55)	<.001	0.36 (0.17 to 0.55)	0.36 (0.17 to 0.54)	<.001
Negative participatory norms	-0.12 (-0.31 to 0.79)	-0.12 (-0.32 to 0.08)	.24	-0.13 (-0.32 to 0.06)	-0.13 (-0.32 to 0.06)	.19
LINK social support	0.20 (-0.15 to 0.41)	0.19 (-0.02 to 0.40)	.07	0.19 (-0.02 to 0.39)	0.18 (-0.02 to 0.38)	.07
LINK brand Perception	0.22 (0.03 to 0.40)	0.21 (0.03 to 0.39)	.02	0.22 (0.04 to 0.40)	0.22 (0.04 to 0.39)	.02

[a]Income recorded in US $20K categories from 1 (<US $20,000) to 6 (>US $100,000).

Table 8. Model parameters for individual-level predictor models for HIV service referral.

Model	N	Listwise deleted	σ_u (95% CI)	σ_e (95% CI)	Rho (95% CI)	LR test of σ_u=0 (*P*)	AIC	BIC
Multiple linear model	451	145	—	—	—	—	1198	1239
Random intercept model	451	145	0.31 (0.10 to 0.97)	1.62 (1.50 to 1.76)	0.03 (0.00 to 0.21)	.008	1196	1244

Table 9. Multilevel predictor models for HIV service referral.[a]

	Neighborhood variable			Model parameters		
Adjusted Neighborhood-Level Predictor[b]	Coefficient (95% CI)	Standardized Coefficient (95% CI)	*P*	Rho (95% CI)	LR test of σ_u=0 (*P*)	AIC
Number of HIV support services	-0.02 (-0.06 to 0.01)	-0.18 (-0.46 to 0.10)	.20	0.01 (0.00 to 0.27)	.27	1196
HIV prevalence	-0.02 (-0.06 to 0.03)	-0.08 (-0.33 to 0.17)	.53	0.04	.009	1197
Black/African American population (%)	0.02 (0.01 to 0.04	0.35 (0.16 to 0.55)	<.001	0.00	>.99	1191
Age ≥ 25 years (%)	-0.03 (-0.09 to 0.03)	-0.14 (-0.44 to 0.16)	.36	0.02 (0.00 to 0.21)	.11	1197
Male high-school graduation	-0.05 (-0.14 to 0.03)	-0.17 (-0.44 to 0.10)	.22	0.03 (0.00 to 0.16)	.006	1196
Male employment (%)	-0.01 (-0.05 to 0.03)	-0.11 (-0.50 to 0.29)	.60	0.02 (0.00 to 0.27)	.17	1197
Median household income (US $K)	-0.07 (-0.11 to -0.03)	-0.33 (-0.53 to 0.14)	<.001	0.00	>.99	1193
Vacant homes (%)	0.08 (0.03 to 0.13)	0.31 (0.12 to 0.49)	.002	0.00	>.99	1194

[a]Each multi-level model includes a single neighborhood-level variable together with all individual and psychosocial variables (N=451; 145 listwise deleted).

[b]Adjusted for race, income, gender, age, and the 5 psychosocial variables.

Figure 3. Individual predictors of CBI Referral.

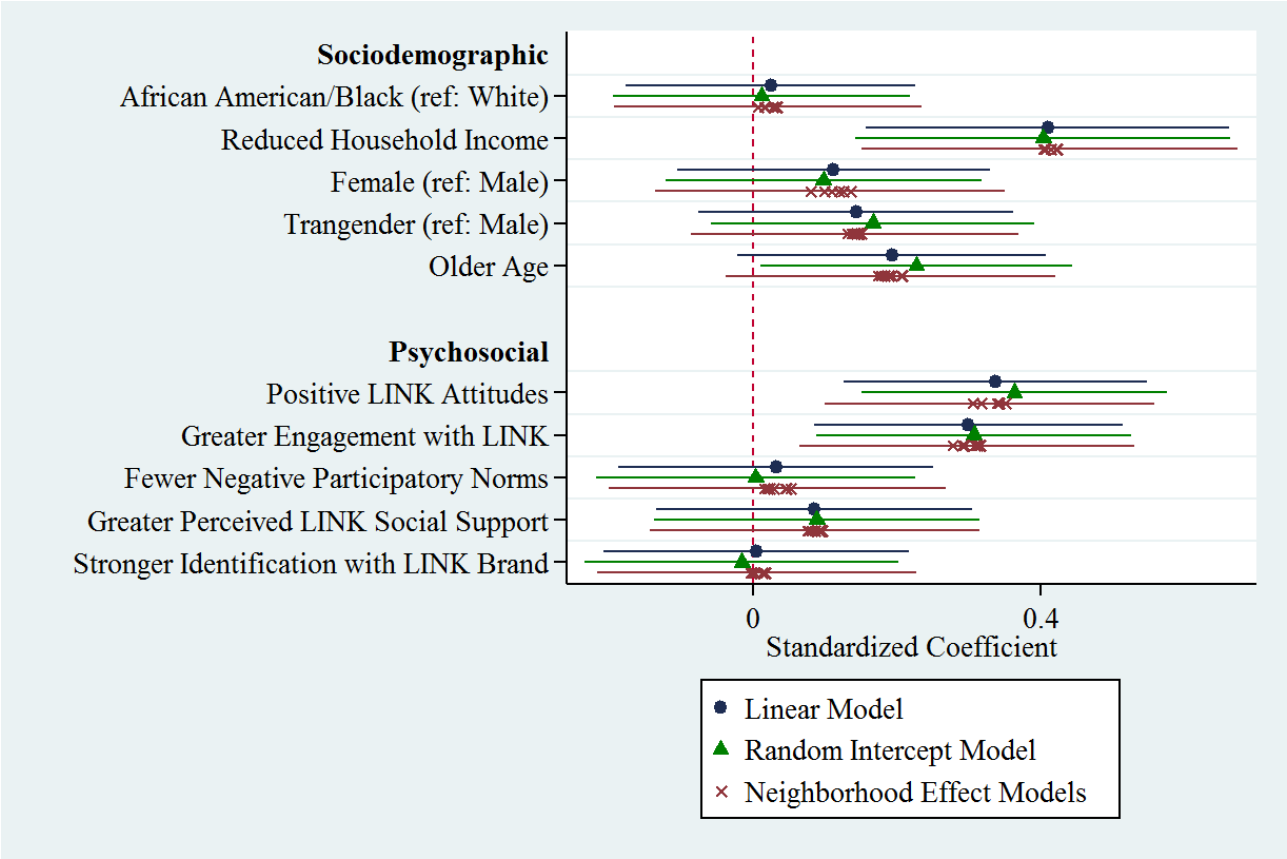
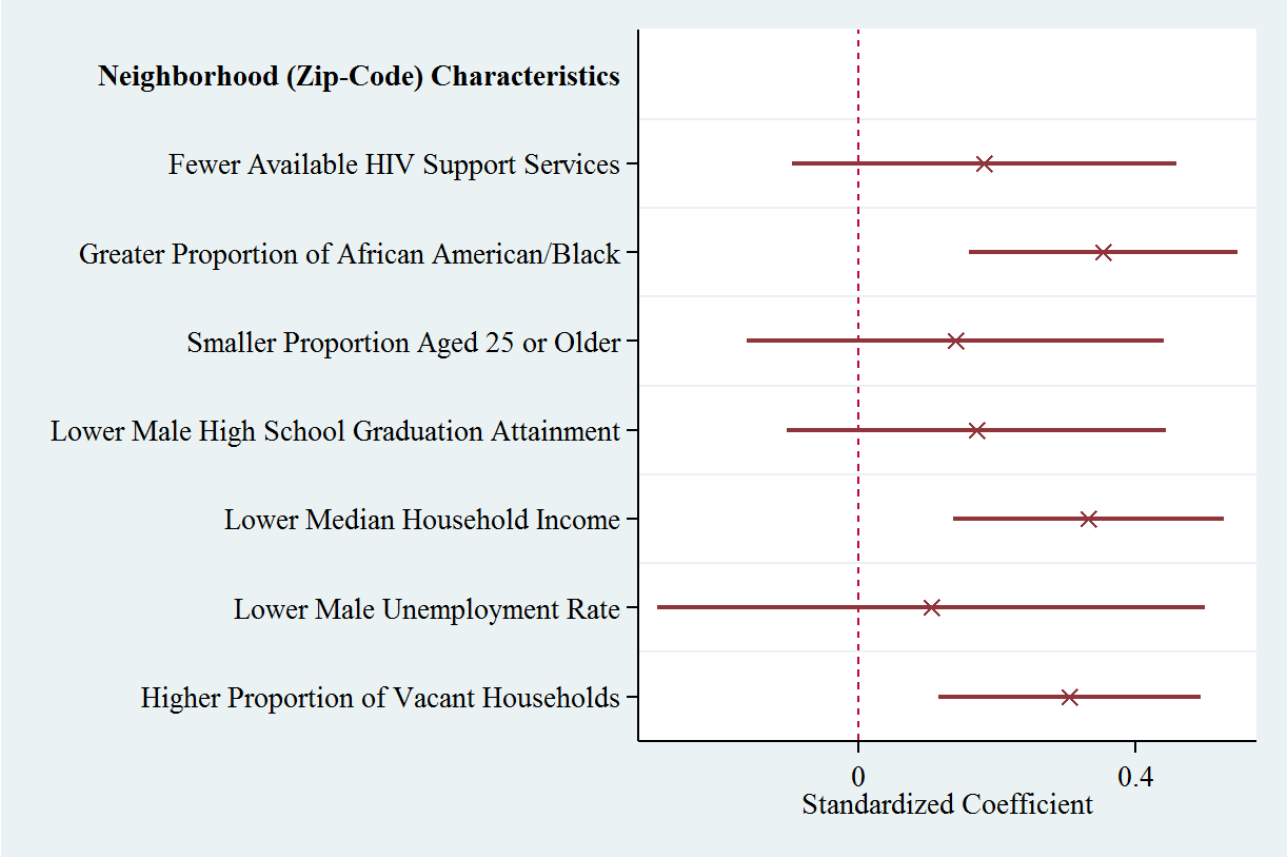


Figure 4. Adjusted neighborhoods predictors of CBI Referral.



Discussion

Summary of Main Findings

This study found that Project LINK participants were extremely positive about their experience with the CBI, and expressed high intentions to use LINK resources to obtain an HIV test and to refer others to the CBI. Both of these outcomes were significantly higher for participants living in target areas compared with those living outside the target and adjoining regions. In addition, for participants in the target and secondary catchment areas, several individual- and neighborhood-level factors were found to be associated with intentions to use LINK resources for HIV testing and with desire to refer others to the CBI. Reduced income, older age, positive attitudes about the CBI, and increased engagement with LINK were all associated with increased desire to use LINK testing resources. On the neighborhood level, intention to test with LINK was associated with reduced availability of support services, reduced HIV prevalence, greater proportion of black/African Americans in the neighborhood, reduced proportion of adults aged 24 years or older, and reduced median household income. Desire to refer others to Project LINK was found to be associated with transgender individuals, positive attitudes about LINK, LINK engagement, identification with the LINK brand, greater proportion of black/African Americans in the neighborhood, reduced median neighborhood household income, and a greater proportion of vacant homes in the neighborhood.

Interpretations and Comparison With Prior Work

These results suggests that Project LINK successfully contributed to HIV testing intentions and HIV service referral among those living in a geographically targeted area characterized by high poverty and HIV prevalence and low availability of HIV services [12]. Greater intention to test and engage others in testing and care referral are critically important outcomes for the realization of continuum of care objectives such as normalized population-level HIV testing, and subsequent referral to and prompt entry into care [9,34].

Living in high-stress neighborhood conditions has been associated with increased risk for HIV/sexually transmitted infection transmission [35], sexual risk behavior, and substance abuse [36]. Recognizing that these factors have contributed to the high HIV prevalence in the LINK-targeted neighborhoods, the CBI partners sought to develop an intervention that was relatively easy to implement through mobile delivery of their existing services to improve community access to diagnosis and care. As a result, those living in the intervention area could select from a menu of 24/7 LINK offerings such as partner-delivered medical and mental health services, domestic violence support groups, family counseling, and individual or group-based “Healthy Love” sexual health and well-being training, all delivered in familiar settings such as homes, schools, churches, community centers, and corner “store front” organizations.

The findings suggest that Project LINK events successfully attracted the enthusiastic participation of the community it served, including those challenged by the environment they

lived in that was characterized by poverty, transportation limitations, and high unemployment. Among the psychosocial factors assessed, we found that positive attitudes about HIV testing and the CBI are critically important for facilitating linkage to testing and referrals to the CBI services.

Notably, LINK drew the participation of black/African American transgender women (transwomen) who represent a highly vulnerable group for HIV transmission [37]. With an estimated 60% of annual HIV infections diagnosed among MSM and transwomen, LINK's culturally sensitive service delivery approach appeared to resonate with the transgender population [38]. The data indicate that transgender individuals showed significantly higher intention to refer others to LINK, even after adjustment for other individual-, psychosocial-, and neighborhood-level covariates. The sample of transgender individuals was small (n=27), so we are cautious of generalizing this result, but we feel that it indicates this CBI model's success in providing a safe and comfortable service environment for those who are marginalized and/or disenfranchised with existing health care options.

Distal neighborhood-level factors played a role in promoting HIV testing intentions and referral to the CBI services [39]. Stronger outcomes were observed among those living in predominately black/African American tracts. We believe these factors can be linked to the messages delivered by LINK partners and others that black/African Americans are disproportionately affected by HIV/AIDS and account for a higher proportion of people living with HIV at every stage from new infection to death [40]. Thus, living in an area with a high case rate among black/African Americans alters individuals' perceptions of their perceived vulnerability to HIV and motivates testing and referral to care and treatment services [41].

We also observed a strong indicator that neighborhood disorder, as evidenced by the density of nearby vacant households, is associated with HIV testing and referral [42]. We assert that this may be indicative of the presence of drug activity including injection drug use and extent of "crack houses" in the neighborhood, both associated with greater concentration of HIV risk behavior and prevalent cases [43-45]. Our study corresponds with previous evidence that intravenous drug users living in lower income areas may be more likely to utilize neighborhood prevention programs than those living in more affluent areas [46]. Thus, our findings evidence the need for targeted community interventions such as LINK in similar high-stress neighborhoods, and reflect the efficacy of LINK in reaching these neighborhoods.

On the individual level, those persons facing the extreme poverty were more likely to intend to obtain HIV testing. There was no difference in testing intentions between men and women, or between white and black/African American participants. Because intention to utilize LINK resources for testing was high on average, especially in the primary target area, this reflects the successful targeting of individuals with the highest need for linkage to care, regardless of race or gender. Although the small number of zip codes in our sample does not allow for effective separation of the neighborhood-level effect across heterogeneous neighborhoods, the large number of significant neighborhood-level factors suggests that community structure plays an important role in the success of targeted HIV care linkage interventions. Among those participants residing in the target area, the reduced availability of local HIV support services resulted in much greater willingness to use LINK for HIV testing. In this respect, we argue that LINK offers a model for reaching historically marginalized populations through its geographically focused, socially compatible service delivery approach [2]. The results also reflect the reality of effectively promoting HIV prevention in communities with considerable challenges; our theoretical orientation was validated by the findings that suggested direct and indirect effects of multiple levels of influence on HIV testing intentions and referral patterns.

The results of this project indicate the importance of including CBIs as options for HIV prevention planning that seeks to increase access to HIV testing and delay the time for linkage to care. Timely linkage to and retention in care is key to ensuring that patients living with HIV reach an undetectable status, which in turns helps to decrease the transmission of new infections. Furthermore, decreasing the waiting time for accessing care also increases the likelihood of people actually beginning therapy, and increasing their chances for improved health while living with HIV. Finally, the outcomes of the project support ensuring that communities and their members are not only engaged when it comes to accessing HIV testing, but are also engaged in the planning, site selection, and coordination of both HIV testing events and activities, as well as strategic planning to ensure linkage to care for any services that will take place within their communities [47]. Thus, CBIs should continue to be used for prevention work in communities, especially when targeting improvements in HIV testing and linkage to care. Linkage strategies should consider incorporating community-level engagement to support timely linkage and retention in care. Bringing services into communities and enabling easy access to neighborhood-based services closer may help reach the immediate linkage-to-care goals of the National AID Strategy and the CDC's High-Impact Prevention programmatic policy.

Limitations

We recognize the limitations associated with self-report, which is susceptible to social desirability bias. In addition, our random sampling approach within venues hosting LINK activities and events is effective for describing associations within the population of attendees [48]. However, this sampling technique together with the locality of the study restricts the ability to generalize results to other neighborhoods, cities, or to nonattendees. Furthermore, privacy concerns limited available address information to only zip-code-level data. While census tracts are generally considered the gold standard for neighborhood-level analysis, zip codes have been successfully used to characterize neighborhood factors in many recent studies in which census-tract-level data are unavailable [49-52]. Although this pilot initiative demonstrated feasibility and acceptability of the concept utilizing attitudinal and intention data, no behavioral data on actual testing were collected. This

is an important next step in examining the efficacy of this project.

Conclusions

Project LINK used a targeted approach to reach marginalized populations that prompted greater intention to engage in HIV testing, care, and referral to community partner organizations serving those in the selected neighborhoods. Our study findings indicate that community/neighborhood and psychosocial factors are critical to future efforts to increase routine HIV testing intentions in underserved areas and enhance efforts for subsequent referral to care. Yet, this project also highlights the need for additional, long-term assessment to evidence the impact of CBIs on care continuum outcomes. The results from this study demonstrate the promise of CBIs to reach individuals within a 1st year of diagnosis for subsequent improvements in HIV-related health outcomes.

Acknowledgments

The authors wish to thank Jamey Rousey and Tene Traylor for their support of this project. Special thanks to the staff of AID Atlanta, the Atlanta Harm Reduction Coalition (Mona Bennett), the Center for Black Women's Wellness (Jamea Dorsey), Positive Impact (Paul Plate), and SisterLove, Inc and all of the community residents and community leaders, Don Speaks and Makeda Johnson.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Results of multilevel models for HIV testing and CBI referral.

[DOCX File, 69KB - publichealth_v1i2e16_app1.docx]

References

1. Salam RA, Haroon S, Ahmed HH, Das JK, Bhutta ZA. Impact of community-based interventions on HIV knowledge, attitudes, and transmission. Infect Dis Poverty 2014;3:26 [FREE Full text] [doi: 10.1186/2049-9957-3-26] [Medline: 25126420]
2. 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]
3. Chen NE, Gallant JE, Page KR. A systematic review of HIV/AIDS survival and delayed diagnosis among Hispanics in the United States. J Immigr Minor Health 2012 Feb;14(1):65-81 [FREE Full text] [doi: 10.1007/s10903-011-9497-y] [Medline: 21773882]
4. MacKellar DA, Valleroy LA, Secura GM, Behel S, Bingham T, Celentano DD, Young Men's Survey Study Group. Unrecognized HIV infection, risk behaviors, and perceptions of risk among young men who have sex with men: Opportunities for advancing HIV prevention in the third decade of HIV/AIDS. J Acquir Immune Defic Syndr 2005 Apr 15;38(5):603-614. [Medline: 15793373]
5. Torian LV, Wiewel EW, Liu K, Sackoff JE, Frieden TR. Risk factors for delayed initiation of medical care after diagnosis of human immunodeficiency virus. Arch Intern Med 2008 Jun 9;168(11):1181-1187. [doi: 10.1001/archinte.168.11.1181] [Medline: 18541826]
6. Tripathi A, Gardner LI, Ogbuanu I, Youmans E, Stephens T, Gibson JJ, et al. Predictors of time to enter medical care after a new HIV diagnosis: A statewide population-based study. AIDS Care 2011 Nov;23(11):1366-1373. [doi: 10.1080/09540121.2011.565032] [Medline: 22022847]
7. Centers for Disease Control and Prevention. HIV Surveillance Supplemental Report. 2015. Monitoring selected national HIV prevention and care objectives by using HIV surveillance data—United States and 6 dependent areas—2013 URL: http://www.cdc.gov/hiv/pdf/library/reports/surveillance/cdc-hiv-surveillancereport_vol20_no2.pdf [accessed 2015-07-14] [WebCite Cache ID 6a0mJZcYo]
8. Anthony MN, Gardner L, Marks G, Anderson-Mahoney P, Metsch LR, Valverde EE, Antiretroviral TreatmentAccess Study (ARTAS) Study Group. Factors associated with use of HIV primary care among persons recently diagnosed with HIV: Examination of variables from the behavioural model of health-care utilization. AIDS Care 2007 Feb;19(2):195-202. [doi: 10.1080/09540120600966182] [Medline: 17364398]
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. Denning P, DiNenno E. Communities in Crisis: Is There a Generalized HIV Epidemic in Impoverished Urban Areas of the United States? Who's At Risk for HIV. Atlanta, GA: Centers for Disease Control and Prevention; 2013. URL: http://www.cdc.gov/hiv/pdf/statistics_poverty_poster.pdf [accessed 2015-05-06] [WebCite Cache ID 6YKAFz0Sh]
11. Shepard CW, Gortakowski HW, Nasrallah H, Cutler BH, Begier EM. Using GIS-based density maps of HIV surveillance data to identify previously unrecognized geographic foci of HIV burden in an urban epidemic. Public Health Rep 2011;126(5):741-749 [FREE Full text] [Medline: 21886335]
12. Hixson BA, Omer SB, del RC, Frew PM. Spatial clustering of HIV prevalence in Atlanta, Georgia and population characteristics associated with case concentrations. J Urban Health 2011 Feb;88(1):129-141 [FREE Full text] [doi: 10.1007/s11524-010-9510-0] [Medline: 21249526]
13. Fulcher C, Kaukinen C. Mapping and visualizing the location HIV service providers: An exploratory spatial analysis of Toronto neighborhoods. AIDS Care 2005 Apr;17(3):386-396. [doi: 10.1080/09540120512331314312] [Medline: 15832887]
14. Oppong JR, Tiwari C, Ruckthongsook W, Huddleston J, Arbona S. Mapping late testers for HIV in Texas. Health Place 2012 May;18(3):568-575. [doi: 10.1016/j.healthplace.2012.01.008] [Medline: 22356835]
15. Eberhart MG, Yehia BR, Hillier A, Voytek CD, Blank MB, Frank I, et al. Behind the cascade: Analyzing spatial patterns along the HIV care continuum. J Acquir Immune Defic Syndr 2013 Nov 1;64 Suppl 1:S42-S51 [FREE Full text] [doi: 10.1097/QAI.0b013e3182a90112] [Medline: 24126447]
16. Philbin MM, Tanner AE, DuVal A, Ellen JM, Xu J, Kapogiannis B, Adolescent Trials Network for HIV/AIDS Interventions. Factors affecting linkage to care and engagement in care for newly diagnosed HIV-positive adolescents within fifteen adolescent medicine clinics in the United States. AIDS Behav 2014 Aug;18(8):1501-1510 [FREE Full text] [doi: 10.1007/s10461-013-0650-6] [Medline: 24682848]
17. Krieger N. Theories for social epidemiology in the 21st century: An ecosocial perspective. Int J Epidemiol 2001 Aug;30(4):668-677 [FREE Full text] [Medline: 11511581]
18. Parker EA, Chung LK, Israel BA, Reyes A, Wilkins D. Community organizing network for environmental health: Using a community health development approach to increase community capacity around reduction of environmental triggers. J Prim Prev 2010 Apr;31(1-2):41-58 [FREE Full text] [doi: 10.1007/s10935-010-0207-7] [Medline: 20306137]
19. Schultz Annette SH, Temple B, Gibbons C, Preston J, Ronson G. Listening to those who are living with HIV and tobacco dependence and exploring their health care context. J Assoc Nurses AIDS Care 2014;25(1):46-59. [doi: 10.1016/j.jana.2013.01.003] [Medline: 23582577]
20. Frew PM, Archibald M, Hixson B, del RC. Socioecological influences on community involvement in HIV vaccine research. Vaccine 2011 Aug 18;29(36):6136-6143 [FREE Full text] [doi: 10.1016/j.vaccine.2011.06.082] [Medline: 21722689]
21. High-Impact HIV Prevention: CDC's Approach to Reducing HIV Infections in the United States. Atlanta, GA: Centers for Disease Control and Prevention; 2011. URL: http://www.cdc.gov/hiv/pdf/policies_NHPC_Booklet.pdf [accessed 2015-05-06] [WebCite Cache ID 6YKBnaTh0]
22. Painter TM, Herbst JH, Diallo DD, White LD, Centers for Disease Control and Prevention (CDC). Community-based program to prevent HIV/STD infection among heterosexual black women. MMWR Surveill Summ 2014 Apr 18;63 Suppl 1:15-20 [FREE Full text] [Medline: 24743662]
23. Diallo DD, Moore TW, Ngalame PM, White LD, Herbst JH, Painter TM. Efficacy of a single-session HIV prevention intervention for black women: A group randomized controlled trial. AIDS Behav 2010 Jun;14(3):518-529. [doi: 10.1007/s10461-010-9672-5] [Medline: 20135214]
24. Maher L, Iversen J, Kaldor J. Measuring effectiveness of needle and syringe exchange programs for prevention of HIV among injecting drug users: Response to Amundsen. Addiction 2006 Dec;101(12):1834-1836. [doi: 10.1111/j.1360-0443.2006.01666.x] [Medline: 17156184]
25. Amundsen EJ. Measuring effectiveness of needle and syringe exchange programmes for prevention of HIV among injecting drug users. Addiction 2006 Jul;101(7):911-912. [doi: 10.1111/j.1360-0443.2006.01519.x] [Medline: 16771880]
26. Holstad MM, Essien JE, Ekong E, Higgins M, Teplinskiy I, Adewuyi MF. Motivational groups support adherence to antiretroviral therapy and use of risk reduction behaviors in HIV positive Nigerian women: A pilot study. Afr J Reprod Health 2012 Sep;16(3):14-27 [FREE Full text] [Medline: 23437496]
27. Rothenberg KH, Paskey SJ, Reuland MM, Zimmerman SI, North RL. Domestic violence and partner notification: Implications for treatment and counseling of women with HIV. J Am Med Womens Assoc 1995;50(3-4):87-93. [Medline: 7657955]
28. Frew PM, Hou SI, Davis M, Chan K, Horton T, Shuster J, et al. The likelihood of participation in clinical trials can be measured: The Clinical Research Involvement Scales. J Clin Epidemiol 2010 Oct;63(10):1110-1117 [FREE Full text] [doi: 10.1016/j.jclinepi.2009.12.002] [Medline: 20303711]
29. Pawinski R. Community attitudes to HIV/AIDS. S Afr Med J 2001 Jun;91(6):448. [Medline: 11455701]
30. Miner MH, Peterson JL, Welles SL, Jacoby SM, Rosser BR. How do social norms impact HIV sexual risk behavior in HIV-positive men who have sex with men?: Multiple mediator effects. J Health Psychol 2009 Sep;14(6):761-770 [FREE Full text] [doi: 10.1177/1359105309338976] [Medline: 19687113]
31. Kumar S, Liu M, Hwang S. A multifaceted comparison of ArcGIS and MapMarker for automated geocoding. Geospat Health 2012 Nov;7(1):145-151. [doi: 10.4081/gh.2012.113] [Medline: 23242688]

32. Census 2000. Washington, DC: United States Census Bureau; 2000. URL: http://factfinder.census.gov/faces/nav/jsf/pages/index.xhtml [accessed 2015-05-06] [WebCite Cache ID 6YKCQzdFd]
33. Boston RC, Sumner AE. STATA: A statistical analysis system for examining biomedical data. Adv Exp Med Biol 2003;537:353-369. [Medline: 14995047]
34. Gardner EM, Daniloff E, Thrun MW, Reirden DH, Davidson AJ, Johnson SC, et al. Initial linkage and subsequent retention in HIV care for a newly diagnosed HIV-infected cohort in Denver, Colorado. J Int Assoc Provid AIDS Care 2013;12(6):384-390. [doi: 10.1177/2325957413500532] [Medline: 23962912]
35. Kerr JC, Valois RF, Siddiqi A, Vanable P, Carey MP, DiClemente RJ, et al. Neighborhood condition and geographic locale in assessing HIV/STI risk among African American adolescents. AIDS Behav 2015 Jun;19(6):1005-1013. [doi: 10.1007/s10461-014-0868-y] [Medline: 25108404]
36. Bowleg L, Neilands TB, Tabb LP, Burkholder GJ, Malebranche DJ, Tschann JM. Neighborhood context and Black heterosexual men's sexual HIV risk behaviors. AIDS Behav 2014 Nov;18(11):2207-2218. [doi: 10.1007/s10461-014-0803-2] [Medline: 24906531]
37. Garofalo R, Deleon J, Osmer E, Doll M, Harper GW. Overlooked, misunderstood and at-risk: Exploring the lives and HIV risk of ethnic minority male-to-female transgender youth. J Adolesc Health 2006 Mar;38(3):230-236. [doi: 10.1016/j.jadohealth.2005.03.023] [Medline: 16488820]
38. Centers for Disease Control and Prevention. Centers for Disease Control and Prevention. Atlanta, GA: Centers for Disease Control and Prevention; 2013. HIV in the US: At a Glance URL: http://www.cdc.gov/hiv/pdf/statistics_basics_factsheet.pdf [accessed 2015-09-08] [WebCite Cache ID 6bOWjkyqK]
39. Raymond HF, Chen Y, Syme SL, Catalano R, Hutson MA, McFarland W. The role of individual and neighborhood factors: HIV acquisition risk among high-risk populations in San Francisco. AIDS Behav 2014 Feb;18(2):346-356. [doi: 10.1007/s10461-013-0508-y] [Medline: 23760633]
40. Centers for Disease Control and Prevention. HIV among African-Americans. Atlanta, GA: Centers for Disease Control and Prevention; 2013. URL: http://www.cdc.gov/hiv/pdf/risk_HIV_AAA.pdf [accessed 2015-09-08] [WebCite Cache ID 6bOWmVAn3]
41. Axelrad JE, Mimiaga MJ, Grasso C, Mayer KH. Trends in the spectrum of engagement in HIV care and subsequent clinical outcomes among men who have sex with men (MSM) at a Boston community health center. AIDS Patient Care STDS 2013 May;27(5):287-296 [FREE Full text] [doi: 10.1089/apc.2012.0471] [Medline: 23651106]
42. McCurdy SA, Ross MW, Kilonzo GP, Leshabari MT, Williams ML. HIV/AIDS and injection drug use in the neighborhoods of Dar es Salaam, Tanzania. Drug Alcohol Depend 2006 Apr;82 Suppl 1:S23-S27. [Medline: 16769441]
43. Latkin CA, Curry AD, Hua W, Davey MA. Direct and indirect associations of neighborhood disorder with drug use and high-risk sexual partners. Am J Prev Med 2007 Jun;32(6 Suppl):S234-S241 [FREE Full text] [doi: 10.1016/j.amepre.2007.02.023] [Medline: 17543716]
44. Latkin CA, German D, Hua W, Curry AD. Individual-level influences on perceptions of neighborhood disorder: A multilevel analysis. J Community Psychol 2009 Jan 1;37(1):122-133 [FREE Full text] [doi: 10.1002/jcop.20284] [Medline: 20027234]
45. Latkin CA, Williams CT, Wang J, Curry AD. Neighborhood social disorder as a determinant of drug injection behaviors: A structural equation modeling approach. Health Psychol 2005 Jan;24(1):96-100. [doi: 10.1037/0278-6133.24.1.96] [Medline: 15631567]
46. Buchanan D, Shaw S, Teng W, Hiser P, Singer M. Neighborhood differences in patterns of syringe access, use, and discard among injection drug users: Implications for HIV outreach and prevention education. J Urban Health 2003 Sep;80(3):438-454 [FREE Full text] [doi: 10.1093/jurban/jtg050] [Medline: 12930882]
47. Centers for Disease Control and Prevention. High-Impact HIV Prevention: CDC's Approach to Reducing HIV Infections in the United States. Atlanta, GA: Centers for Disease Control and Prevention; 2011. URL: http://www.cdc.gov/hiv/pdf/policies_NHPC_Booklet.pdf [accessed 2015-05-06] [WebCite Cache ID 6YKBnaTh0]
48. Muhib FB, Lin LS, Stueve A, Miller RL, Ford WL, Johnson WD, Community Intervention Trial for Youth Study Team. A venue-based method for sampling hard-to-reach populations. Public Health Rep 2001;116 Suppl 1:216-222 [FREE Full text] [Medline: 11889287]
49. Brumback BA, Zheng HW, Dailey AB. Adjusting for confounding by neighborhood using generalized linear mixed models and complex survey data. Stat Med 2013 Apr 15;32(8):1313-1324. [doi: 10.1002/sim.5624] [Medline: 22975990]
50. Lim J, Ashing-Giwa KT. Examining the effect of minority status and neighborhood characteristics on cervical cancer survival outcomes. Gynecol Oncol 2011 Apr;121(1):87-93. [doi: 10.1016/j.ygyno.2010.11.041] [Medline: 21183210]
51. Sathyanarayanan S, Brooks AJ, Hagen SE, Edington DW. Multilevel analysis of the physical health perception of employees: Community and individual factors. Am J Health Promot 2012;26(5):e126-e136. [doi: 10.4278/ajhp.110316-QUAL-120] [Medline: 22548431]
52. Walker DR, Inglese GW, Sloand JA, Just PM. Dialysis facility and patient characteristics associated with utilization of home dialysis. Clin J Am Soc Nephrol 2010 Sep;5(9):1649-1654 [FREE Full text] [doi: 10.2215/CJN.00080110] [Medline: 20634324]

Abbreviations

AHRC: Atlanta Harm Reduction Coalition
AIDS: acquired immune deficiency syndrome
ANOVA: analysis of variance
CBI: Community-based intervention
CBWW: The Center for Black Women's Wellness
CDC: Centers for Disease Control and Prevention
HIV: human immunodeficiency virus
MCAR: Missing Completely at Random
MSM: men who have sex with men

Edited by G Eysenbach; submitted 31.05.15; peer-reviewed by J Weiner, D Cordova; comments to author 02.07.15; revised version received 14.07.15; accepted 29.07.15; published 08.10.15.

Please cite as:
Frew PM, Archibald M, Schamel J, Saint-Victor D, Fox E, Smith-Bankhead N, Diallo DD, Holstad MM, del Rio C
An Integrated Service Delivery Model to Identify Persons Living with HIV and to Provide Linkage to HIV Treatment and Care in Prioritized Neighborhoods: A Geotargeted, Program Outcome Study
JMIR Public Health Surveill 2015;1(2):e16
URL: http://publichealth.jmir.org/2015/2/e16/
doi:10.2196/publichealth.4675
PMID:27227134

Original Paper

Building a Mobile HIV Prevention App for Men Who Have Sex With Men: An Iterative and Community-Driven Process

Tamar Goldenberg[1], MPH; Sarah J McDougal[2], MPH; Patrick S Sullivan[3], DVM, PhD; Joanne D Stekler[2], PhD; Rob Stephenson[1], PhD

[1]School of Nursing, Department of Health Behavior and Biological Sciences and the Center for Sexuality and Health Disparities, University of Michigan, Ann Arbor, MI, United States

[2]Department of Medicine, Division of Allergy and Infectious Diseases, University of Washington, Seattle, WA, United States

[3]Rollins School of Public Health, Department of Epidemiology, Emory University, Atlanta, GA, United States

Corresponding Author:
Tamar Goldenberg, MPH
School of Nursing
Department of Health Behavior and Biological Sciences and the Center for Sexuality and Health Disparities
University of Michigan
400 North Ingalls Building
Ann Arbor, MI,
United States
Phone: 1 734 647 8826
Fax: 1 734 763 7259
Email: tamargol@umich.edu

Abstract

Background: Gay, bisexual, and other men who have sex with men (MSM) account for a disproportionate burden of new HIV infections in the United States. Mobile technology presents an opportunity for innovative interventions for HIV prevention. Some HIV prevention apps currently exist; however, it is challenging to encourage users to download these apps and use them regularly. An iterative research process that centers on the community's needs and preferences may increase the uptake, adherence, and ultimate effectiveness of mobile apps for HIV prevention.

Objective: The aim of this paper is to provide a case study to illustrate how an iterative community approach to a mobile HIV prevention app can lead to changes in app content to appropriately address the needs and the desires of the target community.

Methods: In this three-phase study, we conducted focus group discussions (FGDs) with MSM and HIV testing counselors in Atlanta, Seattle, and US rural regions to learn preferences for building a mobile HIV prevention app. We used data from these groups to build a beta version of the app and theater tested it in additional FGDs. A thematic data analysis examined how this approach addressed preferences and concerns expressed by the participants.

Results: There was an increased willingness to use the app during theater testing than during the first phase of FGDs. Many concerns that were identified in phase one (eg, disagreements about reminders for HIV testing, concerns about app privacy) were considered in building the beta version. Participants perceived these features as strengths during theater testing. However, some disagreements were still present, especially regarding the tone and language of the app.

Conclusions: These findings highlight the benefits of using an interactive and community-driven process to collect data on app preferences when building a mobile HIV prevention app. Through this process, we learned how to be inclusive of the larger MSM population without marginalizing some app users. Though some issues in phase one were able to be addressed, disagreements still occurred in theater testing. If the app is going to address a large and diverse risk group, we cannot include niche functionality that may offend some of the target population.

(JMIR Public Health Surveill 2015;1(2):e18) doi:10.2196/publichealth.4449

KEYWORDS

HIV; AIDS; MSM; mobile app; prevention; community

Introduction

In 2011, gay, bisexual, and other men who have sex with men (MSM) accounted for 62% of new HIV infections in the United States, despite comprising only 2% of the population [1]. To increase identification of new HIV infections and linkage to HIV treatment among MSM, it is recommended that MSM test for HIV at least three to four times per year [2]. However, despite current HIV prevention efforts, most MSM do not test that frequently, with only 20% of MSM testing at least three times per year [3]. This gap identifies a need for new and innovative HIV testing and HIV prevention solutions.

One possible opportunity for innovative HIV prevention is the use of Internet-based interventions and mHealth (the use of mobile phones for medical and public health-supported interventions) [4-9]. mHealth HIV interventions are becoming increasingly popular and include interventions that use mobile text messaging and mobile phone apps [4,10-24]. A study by Muessig et al identified 55 unique mobile apps that address HIV prevention or HIV care, but these apps were not frequently downloaded or highly rated by their users [21]. Even though mobile apps for HIV intervention are a popular platform for developers, it is challenging to encourage app users to download these apps and use them regularly. In order to address this, Muessig et al suggest that prior to building an app, developers should use an iterative data collection process to obtain input from the target audience about app preferences and app evaluation [21].

When disseminating research-based HIV prevention interventions to communities, a disconnect between the research environment and the community can reduce the effectiveness of the intervention and lead to underutilization [25]. Community-based and community-centered approaches towards building research-based interventions can help address the gap between science and practice by ensuring that the interventions are appropriate and driven by the community's needs and desires [26]. When building mHealth interventions, this means interventions should be developed from community-identified needs [4]. An iterative research process that centers on the community's needs and preferences (as suggested by Muessig et al) [21] may increase the uptake, adherence, and ultimate effectiveness of mobile apps for HIV prevention. In this study, we outline the use of a community-driven approach to gather data from MSM in order to build an HIV prevention mobile app. This paper describes how the community-driven approach was used to develop an app that was reflective of the reported needs and desires of the community. We outline an iterative app development process in which multiple rounds of interaction with the target community are used to inform and refine the app content and look. We previously published a report of findings from one round of data collection from focus group discussions (FGDs) with MSM addressing men's preferences for using a mobile HIV prevention app [27]. In the current paper, we build on our previously published work to focus specifically on how multiple rounds of data collection and interaction that constituted the community-driven process allowed us to build a mobile HIV prevention app for MSM that was reflective of their stated needs and desires. The current paper focuses on data from two rounds of data collection, focusing more on the second round of FGDs to illustrate how participant perceptions of the app intervention shifted (or did not shift) after the intervention went through the additional round of app content building. The overall purpose of this paper is to provide a case study to illustrate how an iterative approach to a mobile HIV prevention app can lead to changes in content to appropriately address the preferences of the target community. To achieve this, we present themes that emerged from the data and examine how discussions of these themes shifted at different phases of the study process and how this ultimately lead to the creation of an app that was more likely to be used by the target audience.

Methods

Ethics

This study was approved by Emory University's Institutional Review Board. In this three-phase study, we used FGDs to collect formative data and theater test the app.

Study Population and Recruitment

Methods for recruitment and a description of the FGDs with MSM during phase one have been previously described [27]. We conducted research with three populations: (1) MSM in Atlanta, Seattle, and rural US regions; (2) HIV testing counselors in Atlanta and Seattle; and (3) key informants including primary care providers as well as key stakeholders at community-based organizations, health departments, and other government agencies at local, county, state, and federal levels. From August to December 2013, we recruited MSM using flyers and Facebook advertisements. Flyers were posted in a variety of venues in Atlanta and Seattle (eg, restaurants, bars, coffee shops, gyms) that MSM are known to frequent. Facebook advertisements targeted men living in Atlanta, Seattle, and rural US regions who reported being interested in men in their profiles. In Atlanta, men recruited through Facebook have been reported to be behaviorally comparable to men recruited through other venues [28]. Rural locations were determined by zip codes and defined as geographical areas with population densities of less than 1000 people per square mile using data from the US Census Bureau [29]. The flyers and advertisements provided a link to an online screening survey through SurveyGizmo (Widgix LLC) to determine eligibility for the study. Participants meeting eligibility criteria for MSM in all three geographical regions were self-identified gay or bisexual men aged 18 years or older who owned or had ever owned a smartphone and had never had a positive HIV test. HIV testing counselors were recruited through flyers sent to organizations, clinics, and health departments in Atlanta and Seattle where HIV testing is performed. HIV testing counselors met eligibility criteria if they were aged 18 years or older, provided HIV testing and counseling services to MSM in Atlanta or Seattle, and owned or had ever owned a smartphone. All participants who completed the online survey and were eligible and interested provided email addresses and phone numbers and were later contacted to participate in an FGD. Some MSM were asked to participate in only one FGD (either in phase one or phase three), and others were asked to participate in both; these participants were randomly selected. Key informants were strategically chosen

to provide additional insight into preferences of the users and future collaborations for a smartphone-based HIV prevention app.

Study Procedures

Overview

Other app development studies have used a variety of methods to examine and evaluate mHealth interventions (eg, pre-post test design, interrupted time-series design, randomized controlled testing) [30]. In this study, we applied a simple iterative qualitative approach using FGDs and a beta version of the app to collect preliminary data for building an intervention. The number of FGDs conducted among each population in phases one and three of this study is described in Table 1.

Table 1. Outline of focus group discussions.

Focus group discussions		Atlanta	Seattle	Rural	Total
Phase one					
	FGDs with MSM	2	2	1	5
	FGDs with counselors	1	1	0	2
Phase one total		3	3	1	7
Phase three					
	FGDs with new MSM	1	2	1	4
	FGDS with repeat MSM	1	1	0	2
	FGDs with counselors	1	1	0	2
Phase three total		3	4	1	8
Total		6	7	2	15

Phase One: Focus Group Discussions and Key Informant Interviews

For phase one of this study, we conducted FGDs with MSM and HIV testing counselors to get opinions about what should be included in the HIV prevention app, to understand if and how MSM would use the app, and to determine how the app could be incorporated into HIV counseling sessions. We completed four in-person FGDs with MSM (n=28), one online FGD (OFGD) with rural MSM (n=10) [31], two in-person FGDs with HIV testing counselors (n=13), and 14 key informant interviews. The OFGD used a chatroom-based format in Adobe Connect (Adobe Systems Incorporated), a real-time Web-based meeting client. All FGDs lasted approximately 1.5 hours and were conducted by two trained facilitators (one in Atlanta and one in Seattle) who were familiar with the goals of the mobile HIV prevention app.

All FGDs addressed MSM's general preferences for apps, HIV testing barriers and facilitators, and ways in which an HIV prevention app could address these barriers and facilitators to increase the frequency of HIV testing among MSM. During FGDs, facilitators walked through six images of screenshots to discuss potential functions for a mobile HIV prevention app. Functionality was described by the facilitator, and participants rated each function, providing feedback on why they felt that function would be useful or not useful. Participants also provided suggestions for how to improve each function and the app overall and identified additional functions that should be included.

We also conducted key informant interviews by phone to determine what is feasible and preferable in building the mobile app intervention. Key informants viewed a design document with the same wireframe images of the app. Feedback addressed feasibility of building the app and assessed key informants' interest in collaborating on building the app.

Phase Two: Building a Beta Version of the HIV Prevention App

During phase two, we partnered with Keymind, a division of Axiom Resource Management Inc, to build a beta version of the app. A preliminary analysis of data from phase one was used to build the beta version using a Web-based interactive platform. The mock-up included six major components: (1) navigation aids and pages for personalizing user registration, profile, and privacy and security settings; (2) an interactive HIV testing plan for assessing user testing preferences; (3) a site locator for finding HIV testing facilities; (4) an event tracker for recording sexual encounters, HIV testing dates, and other information relevant to sexual health; (5) frequently asked questions for providing additional HIV prevention tips; and (6) a point system for collecting app interaction credits and donating small denominations of money to organizations focused on HIV and/or lesbian, gay, bisexual, and transgender equity.

Phase Three: Theater Testing

After completion of the beta version of the app, we conducted FGDs to theater test the app and solicit opinions on functionality of the app and how it could be used by MSM to improve HIV prevention. We conducted six FGDs with MSM (n=34), two in Atlanta, three in Seattle, and one OFGD with rural MSM. Two of the six FGDs were with MSM who had participated in the first round of FGDs, and four were with newly recruited MSM. We also conducted two in-person theater testing FGDs with newly recruited HIV testing counselors (n=9), one in Atlanta and one in Seattle.

Theater testing was conducted by the same facilitators who conducted the first round of FGDs. For these groups, the facilitator went through the interactive Web-based beta version of the app piece by piece and asked participants to provide feedback on what they liked and did not like about each feature. Facilitators used scenarios to present possibilities for how MSM could use the app. Participants provided feedback on how each function could be used, their willingness to use it, and suggestions for improvement. The purpose of theater testing was to refine the content of the app, determine the best way to present content, and better understand participant attitudes and willingness to use the app.

Data Analysis

All in-person FGDs were audio-recorded and transcribed verbatim. OFGDs were automatically downloaded to a readable text file. Key informant interviews were not recorded or transcribed, but detailed notes were used to inform the analysis of transcripts from FGDs. Analysis was conducted using MAXQDA version 10 qualitative data analysis software (Verbi GmbH). We conducted a thematic analysis, examining both inductive and deductive themes within the transcripts. After multiple close readings, we created a preliminary codebook of all salient themes. Provisional definitions were given to each code, and four analysts applied each code to a single transcript. The coded transcripts were merged for comparison, and code definitions were revised based on coding disagreements. This process was repeated until a final codebook was created and all four analysts applied codes consistently. Once the final definitions of the codebook were established, analysts consistently applied the codes to all of the fifteen transcripts from both sets of FGDs. Seven of the fifteen transcripts were double-coded with two analysts each coding the same transcript. Eight of the transcripts were coded by one analyst. Double-coded transcripts were merged and codes were reconciled; differences among coders were resolved by consensus. Data were also coded by functionality, with a separate set of inductive codes being applied to all transcripts from phase one and from theater testing. After multiple purposeful and focused readings of coded text, thick descriptions were created for each theme. The descriptions identified common concepts, patterns, and unique ideas expressed in the FGDs. Themes were analyzed separately based on the FGD phase, participant group (MSM or counselors), and location (Atlanta, Seattle, or rural) and were compared and contrasted between groups.

Results

Descriptive Statistics

We conducted 15 FGDs with 70 MSM and 22 HIV testing counselors. Nine of the 70 MSM participated in both phases of FGDs. Participant demographics are described in Tables 2 and 3.

Building on Phase One Results

This three-phase process produced results that enabled researchers and developers to build a detailed design document outlining the functionality of an HIV prevention smartphone app for MSM. Detailed results from FGDs with MSM from phase one have been previously described [27]; however, other data from phase one have not been previously reported. In the first phase of FGDs, MSM described three categories of functions that the app should include: education, interactive engagement, and social networking. MSM also discussed the importance of the tone and privacy of the app. Counselors stated that the app could be a resource during HIV counseling sessions because it could provide educational information, details for risk behavior assessments, previous test dates, etc. Key informants identified potential benefits of the app including benefits for MSM and HIV prevention organizations. Key informants liked that the app could help promote testing and educate MSM about testing, but they also felt that the app should include information about pre-exposure prophylaxis and nonoccupational postexposure prophylaxis and offer more help creating risk reduction plans. Informants also talked about the benefits the app could have for promoting organizations using a locator and potentially allowing MSM to provide feedback on their testing experiences. These informants also identified which functions would be feasible and which would not. For example, we included a wireframe of a function that would validate test results, but community-based organizations and health departments said they do not have the capacity to validate test results on a mobile app. Key informants expressed interest in collaborating and promoting the app.

We used these data from phase one to inform the beta version of the app, resulting in an increased willingness of MSM to use the app during theater testing. Some concerns that were discussed in phase one were addressed in the beta version (eg, disagreements about HIV testing reminders, concerns about app privacy), while others still existed (eg, over functionality, using friendly versus clinical language). We use examples of three app functions (HIV testing reminders, privacy settings, and sex diaries) to explain how the feedback changed throughout the study process. We then examine participant desires for personalizing the app and their willingness to use it.

Table 2. MSM participant demographics and HIV testing behaviors.

		Atlanta n=26	Seattle n=26	Rural n=16	Total n=70
Age, years, mean (range)		32.2 (23-53)	40.9 (19-67)	30.8 (19-48)	35.3 (19-67)
Race, n (%) [a]					
	Non-Hispanic white/Caucasian	16 (62)	21 (78)	14 (88)	51 (74)
	Non-Hispanic black/African American	8 (31)	1 (4)	0 (0)	9 (13)
	Other	2 (8)	5 (19)	2 (13)	9 (13)
Sexual orientation, n (%)					
	Gay/homosexual	24 (92)	26 (93)	14 (88)	64 (91)
	Bisexual	2 (8)	2 (7)	2 (13)	6 (9)
Has had HIV test, n (%)		24 (92)	27 (96)	12 (75)	63 (90)
HIV tests in last 12 months, mean (range)[b]		1.8 (0-4)	1.1 (0-4)	0.7 (0-2)	1.4 (0-4)
Time since last HIV test, n (%) [b]					
	Less than 3 months ago	9 (38)	8 (30)	2 (17)	19 (30)
	3-6 months ago	9 (38)	5 (19)	2 (17)	16 (25)
	6-12 months ago	4 (17)	5 (19)	2 (17)	11 (18)
	More than 1 year ago	2 (8)	4 (15)	5 (42)	11 (18)
	More than 5 years ago	0 (0)	5 (19)	1 (8)	6 (10)
HIV test site (all that apply), n (%) [b]					
	Community-based organization	18 (75)	19 (70.4)	6 (50.0)	43 (68)
	Doctor's office	19 (79)	20 (74.1)	7 (58.3)	46 (73)
	At home	3 (13)	6 (22.2)	1 (8.3)	10 (16)
	Other	3 (13)	6 (22.2)	1 (8.3)	10 (16)

[a]For reporting of race in Seattle (n=27).

[b]Among MSM who have ever been tested for HIV.

Table 3. HIV testing counselor participant demographics.

		Atlanta n=13	Seattle n=9	Total n=22
Age, years, mean (range)		35.9 (23-50)	38.3 (33-50)	37.0 (23-50)
Race, n (%) [a]				
	Non-Hispanic white/ Caucasian	1 (8)	4 (44)	5 (24)
	Non-Hispanic black/ African American	11 (92)	1 (11)	12 (57)
	Other	0 (0)	4 (44)	4 (19)
Gender, n (%)				
	Male	6 (50)	6 (67)	12 (57)
	Female	6 (50)	2 (22)	18 (38)
	Gender queer	0 (0)	1 (11)	1 (5)
Sexual orientation, n (%)				
	Heterosexual	7 (58)	1 (11)	8 (38)
	Gay/homosexual	4 (33)	6 (67)	10 (48)
	Bisexual	1 (8)	1 (11)	2 (10)
	Other	0 (0)	1 (11)	1 (5)
HIV counseling experience, years, mean (range)		2.6 (0.5-9)	3.6 (0.25-12)	3.0 (0.25-12)

[a]Atlanta race demographics are reported on only 12 participants.

The Impact of HIV Testing Reminders on App Privacy

Privacy and security were salient themes in both rounds of FGDs. In phase one, MSM and counselors in all locations expressed concerns about the privacy of the app. This theme was especially salient when discussing the function of having the app provide reminders for HIV testing:

> *It's a matter of privacy with the reminders. I personally don't care if people see my phone when the notification is going to come up. But some people don't want that kind of stuff visible.* [Atlanta FGD 1, round 1]

> *I still don't like the idea of having to explain this [reminder] if my phone is in a visible area… Push notifications are too visible.* [rural OFGD, round 1]

MSM in phase one provided suggestions for how to address these concerns about privacy. One suggestion was to use discreet language to refer to an HIV test, for example, making it too vague for anyone else to understand. Participants also offered suggestions for increasing privacy through how the reminder could be delivered (eg, text message, email, app alerts and banners), but participants within groups disagreed on which would be best, concluding that the best solution would be to offer customization for reminders:

> *I would also agree with what the other comments were around either an email or a text. And I think you should be able to also go in and set your own reminders to say I'd like to get an email every three months to be like hey, you should get tested. I think that would be helpful. I don't like popups. So I find them to be annoying. But I know for other folks, one of the things I would say is that you have to make sure this app is customizable in a lot of different ways because, just around the table, we all have very different preferences around how we use apps.* [Atlanta FGD 2, round 1]

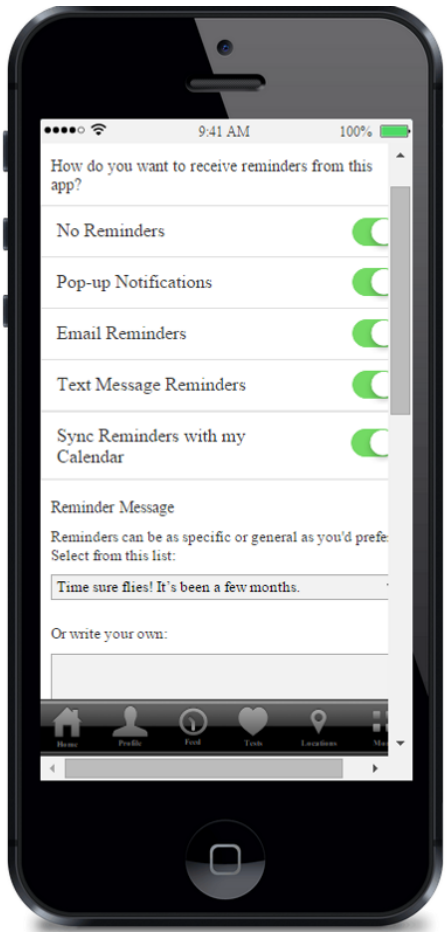
> *I think [reminders are] a good option but overall I think people's preference is so individualized that it would depend on the user the kinds of reminders they prefer.* [rural OFGD, round 1]

In addition to suggesting customization for the mode of delivery for the reminders, participants also suggested the ability to customize the message itself in order to increase privacy. This would enable participants to choose from a list of messages or write in their own discreet message.

When we created the beta version of the app, we applied these suggestions and included customized delivery options for reminders, a list of customized messages to choose from, and the option for the app user to write in his own message (Figure 1). We also provided the option of not receiving reminders. In theater testing, this customization solved the concerns about privacy regarding reminders. Participants who were in both sets of FGDs commented on how their issues from the first round of FGDs had been addressed:

I think it's great. It's personal, we were worried about this last time, about privacy and personally.… I like the fact that there's no reminders there…some people don't want to be reminded. The message is cool. This is exactly what we were worried about like I said last time, and it totally solves that problem. [Atlanta FGD 12, round 2, repeat MSM]

Figure 1. Reminder options.



Additional Privacy Settings

While MSM and counselors focused on the reminders when discussing privacy in the first round of FGDs, they also expressed more general concerns regarding the privacy of the app, including the ability for others using the phone to access the app and the security of the data entered into the app. Many participants wondered what would happen with the data and who would have access to it. In theater testing, we added features to secure privacy, including password protection and three privacy setting options: storing the data locally on the phone, storing the data privately in one's personal cloud, or sharing data anonymously with researchers. Participants really appreciated these options, "Giving people these options will probably cater to everyone" [rural OFGD, round 2]. Nearly all participants stated that they would choose the option of anonymously sharing their data, especially if this would help their community or researchers, help to improve the app, or help app users learn more about patterns of app users in their communities. However, regardless of the reason why a participant would share anonymously, all participants stated MSM who were participating in FGDs for the first time during theater testing expressed that they liked this function and identified which option would be best for them, with different participants choosing different options and describing them as safer in terms of privacy.

that they would only share anonymously if they felt secure in knowing that the data were still protected:

I think it does the most good for society as the larger public health emphasis and all of your data is still protected. [Atlanta FGD 8, round 2, new MSM]

I like the idea of having a general idea of how people are behaving locally, but still want to maintain privacy. [rural OFGD, round 2]

I would probably be more likely to do [option] three [to share data anonymously], only that I think I would get more out of the app that way. [Seattle FGD 9, round 2, new MSM]

Many participants in multiple FGDs advocated so strongly to share data anonymously that they suggested this be the default setting. Alternatively, they suggested forcing app users to address the security settings by having this pop-up on the app when it is first downloaded. Otherwise, participants stated that most app users will simply use the default setting. Being prompted to address the security settings when first downloading

the app was also perceived as increasing the overall feeling of security of the app:

> *If security is the concern it also indicates to them that you thought about security and that this is being addressed upfront rather than like I had to find the security settings* [Seattle FGD 13, round 2, repeat MSM]

Sex Diaries and App Tone

In phase one and phase three of this study, MSM disagreed on the tone of the app. This difference occurred between groups and geographical locations, but also within groups. In the first round of FGDs, some participants wanted more fun and friendly language and functionality, identifying that this would make the app more user-friendly and less judgmental. Other participants identified wanting more clinical language and language and functionality that was more authoritative. This was perceived as increasing the credibility and trustworthiness of the app.

When creating the beta version of the app, we tried to have content and language that addressed both of these needs. Some functionalities (like descriptions of HIV tests) were straightforward, and while we attempted to use simple, easy-to-understand language for this section, it was not written using sexy language. However, other functionality, like sex diaries in the feed (Figure 2), used more fun and sexy language. The sex diaries were meant to help app users track sex partners and experiences to identify risk behaviors, including condom use and substance use during sex. Some MSM in all three locations strongly expressed support for this feature. Participants liked this function because they felt it was fun and might encourage users to be more engaged with the app, possibly encouraging users to try other pieces of the app:

> *It does sound like a certain degree of fun, this app overall, like maybe because you're using this feature of the app, you'll be more likely to use the other features of the app.* [Seattle FGD 9, round 2, new MSM]

Participants also found value in this function because they felt it added accountability by being able to keep track of sexual behavior patterns:

> *I like the idea of tracking your behavior.…I think that that's important when you're creating a testing plan to know when and what did you do and when do I need to go get tested. And then, even after I get tested, have I passed the window period or not, so I'll know if I get tested, do I need to get tested in another couple of weeks, another month or whatever.* [Atlanta FGD 12, round 2, repeat MSM]
>
> *I like it a lot. It makes it feel more like a personal app. Keeping track of things simply adds to the ownership of the whole thing.…People keep a food diary to become healthier. Maybe a sex diary would lead to healthier choices.…And keep one accountable to themselves.* [rural OFGD, round 2]

Many participants in multiple FGDs liked this feature so much they felt we should center the entire app on this function, use the sex diary for marketing, and call it "My Little Black Book."

> *P14: I love the black book idea. I'd market the hell out of this aspect, and use it as tool for HIV testing as the secondary.*
>
> *Moderator: How do others feel about that?*
>
> *P18: But, P14, the whole POINT is HIV testing.*
>
> *P14: I know, but use the fun part of it to get people recording and thinking about their activities and then use that history to encourage testing. …more of a cover on the testing things.*
>
> *P22: That's true. A Black Book app would be a good angle on it. Maybe Black Book could be incorporated into the title?*
>
> *P14: Take the clinical aspect out of it.*
>
> *P19: I feel this is getting a little off-track.*
>
> *P18: Then again, point. See the list of stuff you've done, and be like, "I should get tested."*
>
> *P14: Yea… I'm not suggesting dropping the testing, but the diary aspect puts your history in your face, makes you think about it more.* [rural OFGD, round 2]

Despite the overall positive reaction to this feature in many of the FGDs, not all participants liked this feature. Some, like P19 stated in the OFGD, felt that it took away from the main point of the app.

Even when participants found value in the sex diary function, some participants felt that this function asked for too much from app users and they did not want to put that type of information into their phone:

> *Are you really going to go to the effort of putting in the dirty details?* [rural OFGD, round 2]

Sometimes these concerns were related to app user motivation, but participants also identified privacy as a concern regarding this type of sensitive information:

> *I was the one advocating for blunt use of language but at the same time no one would write that to themselves on the off chance that their mother picked up the phone, but still it's useful information…even for your own personal diary or your own personal use.* [Atlanta FGD 8, round 2, new MSM]
>
> *It's just very personal. It's up to the individual, I mean, I certainly wouldn't put anything like that on my phone, but somebody that wants to get that detailed and it's their own personalization, I guess.* [Atlanta FGD 12, round 2, repeat MSM]

Some participants (especially those in Atlanta) disagreed with this function even further and did not find value in its use. Some of these participants stated that they would delete the app if this were a function on it, even if this function were optional. These participants felt that this function took away from the main point of the app and would encourage MSM to brag to their friends about the number of sex partners they had had:

P1: I'm just kind of uncomfortable now so I don't think I would put that on my phone…

P2: I would not take this app seriously after seeing this.

P1: I would completely walk away, be done… I don't see the purpose of writing [about] the sex…Unless you wanted to take this app…

P2: Unless you wanted to make it a game or something.

P3: Well then does it start to defeat its own purpose once you start turning this into like a super fun, how many things can I list out.

P1: Yeah, you start to look at the game and then it's like…the HIV part becomes, 'Oh by the way go get tested' after I've done all this.

P2: Knowing that, you know, [Name] blah, blah, blah, has his sex diary on his phone who wants to see his sex diary then he passes his phone around the bar. [Atlanta FGD 8, round 2, new MSM]

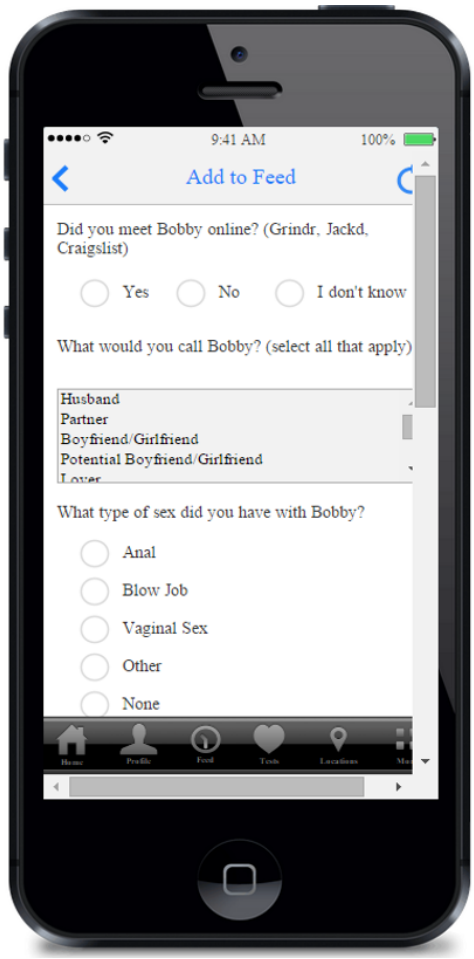
Some participants who disagreed with this function felt that it would have been improved by a more professional tone:

And this may be just a personal preference for me but I would keep the language… extremely clinical… if I were going to use it I'd want it to be something where I could just put the facts down so that I could have a reminder if I needed it. [Atlanta FGD 8, round 2, new MSM]

This disagreement about the sex diary is an example of the tension between making the language and tone of the app overly clinical and over-sexing the language and functionality of the app so that it offends people and deters them from using the app. There was no expressed solution to solve this disagreement, but participants recognized that this tension is a sensitive issue that can make or break an app:

I think there is a fine line between keeping it real and being accessible and trivializing. [Atlanta FGD 8, round 2, new MSM]

Figure 2. Sex diary.



Personalization of the App

Participants discussed the importance of personalizing the app. This was evident in the appreciation of customized features like the reminders and the privacy settings. Participants who approved of the sex diaries also stated that they felt it made the app more personal and added to the ownership of the whole thing. In addition to functionality that personalized the app, participants also stated that the language that the app used was personal:

The personal pronouns lend the user towards a sense of ownership of the whole thing. [rural OFGD, round 2]

> *Something that sort of personalizes, like you've got MY feed, MY test plan using the word my.* [Atlanta FGD 8, round 2, new MSM]

Using titles for functions such as "My Test Plan" helped participants to take ownership over these functions and identified that these were functions that could be customized and personalized to fit each app user's individual needs. Participants liked this type of language so much that they suggested that the title of the app include a personal pronoun to stress the importance of ownership. According to participants, ownership applied to the ability to customize and personalize app functions, but it also contributed to the ownership of one's sexual health and HIV risk through increased self-responsibility.

Willingness to Use the App

Overall, participants felt that the app seemed easy to use and the information was easy to digest. Participants appreciated the simple, straightforward language as well as infographics and suggested including more of these. Participants recognized that men most likely to use or need the app are men who are more sexually active, who have concerns about their HIV risk, or who do not already have a HIV testing plan established.

One of the biggest challenges identified by participants in getting men to use the app is maintaining interest and motivation to use the app:

> *It seems like one of those apps that you download and you play it for about ten minutes after you downloaded it and then just kind of sat there on your phone until you get the notification.* [Atlanta FGD 8, round 2, new MSM]

Even though some participants expressed that this might not be an app that they would use constantly, participants did see the importance for having this app available when they really needed it:

> *But I think too if you do have an accident or you have engaged in high risk behavior…and you're scared and you don't have a plan in place you might turn to this out of curiosity to help build a plan. It would be a good anonymous way to…create a plan and find out as much as you want to find out too. But…I wouldn't play on it every day but again we've all had.* [Atlanta FGD 8, round 2, new MSM]

Participants also offered suggestions for what could help motivate MSM to use the app more regularly. One idea was to incentivize app use. In the beta version, we included incentives in the form of reward points that would contribute to donations for organizations, but participants stated that there should also be incentives that benefit the app user directly:

> *I was trying to think of what would get me or get some people I know to do it, and it might be like, I was wondering if you could do, you know, fifty points and we'll send you a pack of condoms, or a hundred points and you get no cover charge at this club.…I think [the reward points are a] great thing, I don't know if it's going to motivate as many people as something that's actually for them.* [Seattle FGD 10, round 2, new MSM]

In addition to incentivizing, many participants in multiple FGDs discussed app promotion and advertising as an important way to get users to download the app.

Discussion

Principal Findings

These findings highlight how using an interactive and community-centered process to collect data on app preferences is fundamental when building a mobile HIV prevention app. Many of the concerns and problems that were voiced in the first round of FGDs were addressed in the beta version of the app with increased acceptability noted in phase three, especially regarding concerns about privacy. Through this process, we learned about the needs and desires that MSM have for a mobile HIV prevention app and gained insight on what would motivate men to download and use the app.

Through the testing, we learned that if this app-based intervention is going to address a large and diverse risk group, we cannot include niche functionality that may offend some of the target population. Even though some participants loved the sex diaries, others said that they would not use the app at all if it was included, even as an optional function. This app is meant to cater to the larger MSM population, so it needs to include more general functionality that everyone agrees is useful while also being customizable so that each app user can have a personalized experience. In the process of building this app, we learned that personalization and customization can improve many components of the app, especially when there are personalized settings to address different user security needs. This personalization along with interactive functionality allows for the app user to take ownership over the app, making HIV testing plans and other features more catered to the app users' specific needs. According to participants, this act of taking ownership over one's sexual health within the app may also assist men to take ownership of their HIV risk management in other aspects of their lives.

This concept of personalization and ownership is aligned with Bandura's social cognitive theory of self-regulation, which states that self-regulation occurs through self-monitoring, judging one's behaviors in relation to personal and societal standards, and reacting to these judgments [32]. Participants who appreciated the sex diaries expressed the benefits of self-regulation and increased accountability. Participants also recognized other areas of the app that encouraged this type of accountability, such as the use of personal pronouns (eg, My Test Plan). These findings support current recommendations that mHealth interventions should be guided by existing theories of behavior change [33]. HIV prevention apps may benefit from applying this cognitive theory of self-regulation through increased personalization, accountability, and ownership of app functionality.

Limitations

Although recruitment by race is reflective of the larger geographical demographics [34] with more black MSM

participating in Atlanta, we do not have enough racial variation to make any conclusions based on race. Participants represented an older population of MSM with an overall mean age of 35. We do not know how the results would have differed if they were more reflective of the attitudes of younger MSM, and given the increase in HIV among young MSM, this represents a knowledge gap in the current study. Participants in Seattle also had a higher mean age than participants in Atlanta and we do not know if this contributed to any differences between geographical groups. In addition, recruitment of men who identify as gay or bisexual and who identify as seeking other men on Facebook may not represent MSM in general. Including only MSM who own smartphones may also not be representative of MSM in general; however, targeting MSM who owned smartphones allowed us to include participants who would be most likely to use the app. Since these are qualitative findings, we are not trying to generalize study results or quantify differences between groups. However, we were able to understand app preferences using populations in two different US cities as well as rural regions, where population demographics, culture, and HIV prevention efforts vary. OFGDs with rural men were limited to the online environment. All participants needed to have access to a computer with Internet. Furthermore, the facilitator did not have nonverbal cues to assist with probing questions. Still, the OFGDs were useful in capturing insight from a population that we would have otherwise not been able to include in this study. Overall, this study addressed perspectives and attitudes on willingness to use an HIV prevention app and to determine the best approach for functionality and content of the app; however, this study did not address usability testing. While usability testing is important to ensure that the target population will use the app, since this paper describes an early stage of app development, the aim of the FGDs was to determine preferences for content and functionality rather than design and interface. Future testing of the app will address other factors that influence app usability.

Conclusions

Despite these limitations, this study reflects the need for a community-driven approach that includes multiple rounds of data collection and theater testing when developing apps or other mHealth interventions for HIV prevention. Building an HIV prevention app is expensive and requires time and resources. To maximize app uptake and usage, it makes sense to build the best app possible, and the definition of best app should be defined by the community it aims to serve. Through this process, we learned how to be inclusive of the larger MSM population without marginalizing some app users. We also learned how to personalize the app so users take ownership and feel comfortable with its security. This community-driven process increased an overall willingness to use the app and provided important insight into how to build an HIV prevention app that MSM want to use.

Acknowledgments

This research was supported by MAC AIDS Fund and the Emory Center for AIDS Research (P30 AI050409).

Conflicts of Interest

None declared.

References

1. Estimated HIV Incidence in the United States, 2007-2010. Atlanta GA: Centers for Disease Control & Prevention; 2012. HIV Surveillance Supplemental Report URL: http://www.cdc.gov/hiv/pdf/statistics_hssr_vol_17_no_4.pdf [accessed 2015-10-30] [WebCite Cache ID 6cfjrJUcQ]
2. Oster AM, Johnson CH, Le BC. Trends in HIV prevalence and HIV testing among young MSM: five United States Cities, 1994-2011. AIDS Behav 2014;18(3):237-247. [doi: 10.1007/s10461-013-0566-1]
3. Khosropour CM, Sullivan PS. At-home HIV testing of MSM enrolled in an online HIV behavioral risk study. 2011 Presented at: National HIV Prevention Conference; August 14-17 2011; Atlanta.
4. Arya M, Kumar D, Patel S, Street RL, Giordano TP, Viswanath K. Mitigating HIV health disparities: the promise of mobile health for a patient-initiated solution. Am J Public Health 2014 Dec;104(12):2251-2255. [doi: 10.2105/AJPH.2014.302120] [Medline: 25322292]
5. Boulos Maged N Kamel, Wheeler S, Tavares C, Jones R. How smartphones are changing the face of mobile and participatory healthcare: an overview, with example from eCAALYX. Biomed Eng Online 2011;10:24 [FREE Full text] [doi: 10.1186/1475-925X-10-24] [Medline: 21466669]
6. Mobile Health 2012: Half of smartphone owners use their devices to get health information and one-fifth of smartphone owners have health apps. Washington DC: Pew Internet & American Life Project URL: http://www.pewinternet.org/files/old-media/Files/Reports/2012/PIP_MobileHealth2012_FINAL.pdf [accessed 2015-10-30] [WebCite Cache ID 6cfkTL8pF]
7. Terry M. Medical Apps for Smartphones. Telemed J E Health 2010;16(1):17-22. [doi: 10.1089/tmj.2010.9999] [Medline: 20070172]
8. Ybarra ML, Bull SS. Current trends in Internet- and cell phone-based HIV prevention and intervention programs. Curr HIV/AIDS Rep 2007 Dec;4(4):201-207. [Medline: 18366952]

9. Sullivan PS, Grey JA, Simon Rosser BR. Emerging technologies for HIV prevention for MSM: what we have learned, and ways forward. J Acquir Immune Defic Syndr 2013 Jun 1;63 Suppl 1:S102-S107 [FREE Full text] [doi: 10.1097/QAI.0b013e3182949e85] [Medline: 23673879]
10. Catalani C, Philbrick W, Fraser H, Mechael P, Israelski DM. mHealth for HIV Treatment & Prevention: A Systematic Review of the Literature. Open AIDS J 2013;7:17-41 [FREE Full text] [doi: 10.2174/1874613620130812003] [Medline: 24133558]
11. Cornelius JB, Cato M, Lawrence JS, Boyer CB, Lightfoot M. Development and pretesting multimedia HIV-prevention text messages for mobile cell phone delivery. J Assoc Nurses AIDS Care 2011;22(5):407-413 [FREE Full text] [doi: 10.1016/j.jana.2010.11.007] [Medline: 21256053]
12. Cornelius JB, Cato MG, Toth JL, Bard PM, Moore MW, White A. Following the trail of an HIV-prevention Web site enhanced for mobile cell phone text messaging delivery. J Assoc Nurses AIDS Care 2012;23(3):255-259 [FREE Full text] [doi: 10.1016/j.jana.2011.03.002] [Medline: 21550826]
13. Cornelius JB, St Lawrence JS, Howard JC, Shah D, Poka A, McDonald D, et al. Adolescents' perceptions of a mobile cell phone text messaging-enhanced intervention and development of a mobile cell phone-based HIV prevention intervention. J Spec Pediatr Nurs 2012 Jan;17(1):61-69 [FREE Full text] [doi: 10.1111/j.1744-6155.2011.00308.x] [Medline: 22188273]
14. da Costa TM, Barbosa BJ, Gomes e Costa DA, Sigulem D, de Fátima 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]
15. 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]
16. Hardy H, Kumar V, Doros G, Farmer E, Drainoni M, Rybin D, et al. Randomized controlled trial of a personalized cellular phone reminder system to enhance adherence to antiretroviral therapy. AIDS Patient Care STDS 2011 Mar;25(3):153-161 [FREE Full text] [doi: 10.1089/apc.2010.0006] [Medline: 21323532]
17. Horvath T, Azman H, Kennedy GE, Rutherford GW. Mobile phone text messaging for promoting adherence to antiretroviral therapy in patients with HIV infection. Cochrane Database Syst Rev 2012;3. [Medline: 22419345]
18. Juzang I, Fortune T, Black S, Wright E, Bull S. A pilot programme using mobile phones for HIV prevention. J Telemed Telecare 2011;17(3):150-153. [doi: 10.1258/jtt.2010.091107] [Medline: 21270049]
19. Krishnan A, Ferro EG, Weikum D, Vagenas P, Lama JR, Sanchez J, et al. Communication technology use and mHealth acceptance among HIV-infected men who have sex with men in Peru: implications for HIV prevention and treatment. AIDS Care 2015;27(3):273-282. [doi: 10.1080/09540121.2014.963014] [Medline: 25285464]
20. Lester RT, Ritvo P, Mills EJ, Kariri A, Karanja S, Chung MH, et al. Effects of a mobile phone short message service on antiretroviral treatment adherence in Kenya (WelTel Kenya1): a randomised trial. Lancet 2010 Nov 27;376(9755):1838-1845. [doi: 10.1016/S0140-6736(10)61997-6] [Medline: 21071074]
21. Muessig KE, Pike EC, Legrand S, Hightow-Weidman LB. Mobile phone applications for the care and prevention of HIV and other sexually transmitted diseases: a review. J Med Internet Res 2013;15(1):e1 [FREE Full text] [doi: 10.2196/jmir.2301] [Medline: 23291245]
22. Pop-Eleches C, Thirumurthy H, Habyarimana JP, Zivin JG, Goldstein MP, de WD, et al. Mobile phone technologies improve adherence to antiretroviral treatment in a resource-limited setting: a randomized controlled trial of text message reminders. AIDS 2011 Mar 27;25(6):825-834 [FREE Full text] [doi: 10.1097/QAD.0b013e32834380c1] [Medline: 21252632]
23. 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]
24. Tufts KA, Johnson KF, Shepherd JG, Lee JY, Bait Ajzoon MS, Mahan LB, et al. Novel interventions for HIV self-management in African American women: a systematic review of mHealth interventions. J Assoc Nurses AIDS Care 2015;26(2):139-150. [doi: 10.1016/j.jana.2014.08.002] [Medline: 25283352]
25. Kelly JA, Somlai AM, DiFranceisco WJ, Otto-Salaj LL, McAuliffe TL, Hackl KL, et al. Bridging the gap between the science and service of HIV prevention: transferring effective research-based HIV prevention interventions to community AIDS service providers. Am J Public Health 2000 Jul;90(7):1082-1088. [Medline: 10897186]
26. Wandersman A, Florin P. Community interventions and effective prevention. Am Psychol 2003;58(6-7):441-448. [Medline: 12971190]
27. 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.
28. Hernandez-Romieu AC, Sullivan PS, Sanchez TH, Kelley CF, Peterson JL, Del Rio C, et al. The comparability of men who have sex with men recruited from venue-time-space sampling and facebook: a cohort study. JMIR Res Protoc 2014;3(3):e37 [FREE Full text] [doi: 10.2196/resprot.3342] [Medline: 25048694]

29. Womach J. Congressional Research Service Report for Congress. 2005. Agriculture: A Glossary of Terms, Programs, and Laws, 2005 Edition URL: http://digital.library.unt.edu/ark:/67531/metacrs7246/ [accessed 2015-10-30] [WebCite Cache ID 6cfm1CFRu]
30. Kumar S, Nilsen WJ, Abernethy A, Atienza A, Patrick K, Pavel M, et al. Mobile health technology evaluation: the mHealth evidence workshop. Am J Prev Med 2013 Aug;45(2):228-236 [FREE Full text] [doi: 10.1016/j.amepre.2013.03.017] [Medline: 23867031]
31. Stewart K, Williams M. Researching online populations: the use of online focus groups for social research. Qualitative Research 2005 Nov 01;5(4):395-416. [doi: 10.1177/1468794105056916]
32. Bandura A. Social cognitive theory of self-regulation. Organ Behav Hum Decis Process 1991;50(2):248-287.
33. Tomlinson M, Rotheram-Borus MJ, Swartz L, Tsai AC. Scaling up mHealth: where is the evidence? PLoS Med 2013;10(2):e1001382 [FREE Full text] [doi: 10.1371/journal.pmed.1001382] [Medline: 23424286]
34. QuickFacts. Washington DC: United States Census Bureau; 2010. URL: http://quickfacts.census.gov/ [accessed 2015-10-30] [WebCite Cache ID 6cfmUd9H3]

Abbreviations

FGD: focus group discussion
MSM: men who have sex with men
OFGD: online focus group discussion

Edited by G Eysenbach; submitted 20.03.15; peer-reviewed by K Horvath, R Schnall; comments to author 16.04.15; revised version received 22.05.15; accepted 22.09.15; published 16.11.15.

Please cite as:
Goldenberg T, McDougal SJ, Sullivan PS, Stekler JD, Stephenson R
Building a Mobile HIV Prevention App for Men Who Have Sex With Men: An Iterative and Community-Driven Process
JMIR Public Health Surveill 2015;1(2):e18
URL: http://publichealth.jmir.org/2015/2/e18/
doi:10.2196/publichealth.4449
PMID:27227136

Protocol

The Tobacco Pack Surveillance System: A Protocol for Assessing Health Warning Compliance, Design Features, and Appeals of Tobacco Packs Sold in Low- and Middle-Income Countries

Katherine Smith[1], PhD; Carmen Washington[1], MPH, MSW; Jennifer Brown[1], MPH; Alison Vadnais[1], MHS; Laura Kroart[1], BA; Jacqueline Ferguson[1], MHS; Joanna Cohen[1], PhD

Institute for Global Tobacco Control, Department of Health, Behavior and Society, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, United States

Corresponding Author:
Katherine Smith, PhD
Institute for Global Tobacco Control
Department of Health, Behavior and Society
Johns Hopkins Bloomberg School of Public Health
Rm 726
624 N. Broadway
Baltimore, MD, 21215
United States
Phone: 1 4105020025
Fax: 1 4109557241
Email: katecsmith@jhu.edu

Abstract

Background: Tobacco remains the world's leading preventable cause of death, with the majority of tobacco-caused deaths occurring in low- and middle-income countries. The first global health treaty, the Framework Convention on Tobacco Control (FCTC), outlines a set of policy initiatives that have been demonstrated as effective in reducing tobacco use. Article 11 of the FCTC focuses on using the tobacco package to communicate tobacco-caused harms; it also seeks to restrict the delivery of misleading information about the product on the pack.

Objective: The objective of this study was to establish a surveillance system for tobacco packs in the 14 low- and middle-income countries with the greatest number of smokers. The Tobacco Pack Surveillance System (TPackSS) monitors whether required health warnings on tobacco packages are being implemented as intended, and identifies pack designs and appeals that might violate or detract from the communication of harm-related information and undermine the impact of a country's tobacco packaging laws. The protocol outlined is intended to be applicable or adaptable for surveillance efforts in other countries.

Methods: Tobacco packs were collected in 14 countries during 2013. The intention was, to the extent possible, to construct a census of "unique" pack presentations available for purchase in each country. The TPackSS team partnered with in-country field staff to implement a standardized protocol for acquiring packs from 36 diverse neighborhoods across three cities in each country. At the time of purchase, data on price and place of acquisition of each pack was recorded. The field staff, according to a standardized protocol, then photographed packs before they were shipped to the United States for coding and archiving.

Results: Each pack was coded for compliance with the country-specific health warning label laws, as well as for key design features of the pack and appeals of the branding elements. The coding protocols were developed based upon prior research, expert opinion, and communication theories. Each pack was coded by two independent coders, with consistency of personnel across the project. We routinely measured intercoder reliability, and only retained variables for which a good level of reliability was achieved. Variables where reliability was too low were not included in final analyses, and any inconsistencies in coding were resolved on a daily basis.

Conclusions: Across the 14 countries, the TPackSS team collected 3307 tobacco packs. We have established a publicly accessible, Internet archive of these packs that is intended for use by the tobacco control policy advocacy and research community.

(JMIR Public Health Surveill 2015;1(2):e8) doi:10.2196/publichealth.4616

KEYWORDS

tobacco products; cigarettes; public health surveillance; health communication; national health policy; marketing; developing countries

Introduction

Policy Approaches to Tackling the Global Tobacco Epidemic

Tobacco use kills six million people each year and remains the leading preventable cause of death around the globe [1]. Almost 80% of tobacco-related deaths occur in low- and middle-income countries [2]. The World Health Organization's Framework Convention on Tobacco Control (FCTC), the first global health treaty, outlines the evidence-based policies that countries should adopt in order to eliminate tobacco use [3]. The surveillance system outlined in this paper pertains to Article 11 of the FCTC, which requires that parties to the FCTC implement effective packaging and labeling measures to increase public awareness of the negative health impacts of tobacco products.

The Cigarette Pack as a Powerful Communication Platform

The cigarette pack is on display not only when cigarettes are purchased, but also each time a smoker retrieves a cigarette from the pack [4]. A person who smokes a pack of cigarettes per day might look at a pack over 7000 times a year [5]. Moreover, it is not only smokers who are exposed to cigarette packs. Packs are in public view much of the time, either in the hands of smokers, left out at a social gathering, or prominently displayed in retail settings [5]. The cigarette pack has long been a key marketing method to attract new smokers and retain current ones [6-8]. As restrictions on tobacco marketing and advertising tighten, so tobacco packs have become an ever more important channel for brand advertising [6].

In addition to promoting smoking and prompting purchase of a specific brand, cigarette packs can also be used to provide important health information to smokers and the wider public. A growing number of countries are implementing health warning labels (HWLs) on cigarette packs, consistent with Article 11 of the FCTC [9]. As of September 2014, 77 countries or jurisdictions had adopted and finalized graphic HWL requirements [10]. HWLs have been associated with increased awareness of smoking risks; reduced appeal of smoking and smoking initiation among youth; increased motivation and intention to quit among smokers; increased cessation behaviors; increased use of cessation resources; increased likelihood that ex-smokers will remain abstinent; and reduced consumption levels among smokers [11-15]. Moreover, HWLs have been found to be a prominent and trusted source of health information for both smokers and nonsmokers [5,16]. Not all health warnings are, however, equally effective. Exposure to health warnings that are larger, placed on the front upper face of the tobacco package, and that contain pictures, are more effective than smaller text-only warnings [5,8,17,18]. While the impact of all warnings has been shown to lessen over time, pictorial warnings are less impacted by the "wear-out" effect [16,17]. Larger warnings allow for more information, such as cessation resources, to be provided to the consumer [15]. Graphic warnings that elicit a strong emotional response have been shown to be better remembered and reduce the urge to smoke [19].

In addition to adding warnings to packs, a number of countries now prohibit descriptors such as "light", "mild", and "low tar" on tobacco packaging on the basis that such terms can deceptively imply that these products are less harmful than other products. In the place of these descriptors, tobacco companies are using proxy terms (such as "smooth"), colors, and design elements to convey (falsely) that some products are less harmful than others [6,9,20-22].

Study Overview

The Institute for Global Tobacco Control (IGTC) at the Johns Hopkins Bloomberg School of Public Health has developed the Tobacco Pack Surveillance System (TPackSS), to monitor whether required health warnings on tobacco packages are being implemented as intended, and to identify pack design features and appeals that might violate or detract from HWLs.

TPackSS was designed to be implemented in the 14 low- and middle-income countries with the greatest number of smokers [1]. The overall goal of the project is to collect one of every unique pack available for sale in each country. The initiative was developed with funding from Bloomberg Philanthropies through the Bloomberg Initiative to Reduce Tobacco Use [23].

Methods

Development of the Tobacco Pack Surveillance System Protocol

The development of the TPackSS protocol began with an exploratory and planning phase in 2012 to identify strategic goals and systematic protocols for establishing a surveillance system for cigarette packs. Based on discussions with 11 key tobacco control informants, a protocol was developed to collect packs as well as to train data collectors, photograph packs, code packs, and construct a searchable, Internet archive. Each expert consulted is affiliated with a leading academic institution or a public sector research or advocacy organization focusing on tobacco control issues. None of the consultants for this project have ties to the tobacco industry.

Partnering With In-Country Data Collection Teams

In each country, IGTC collaborated with an in-country agency (market research firm; academic research group; nongovernmental organization, NGO; government-related agency; or independent consultant) to conduct field activities (see Table 1).

Table 1. In-country agency information and data collection dates.

Country	Agency type	Dates of collection
Bangladesh	NGO	September-October 2013
Brazil	NGO	January 2013
China	Academic research group	November-December 2013
Egypt	NGO	November-December 2013
India	Market research firm	October 2013
Indonesia	NGO	November 2013
Mexico	Government-related agency	July-August 2013
Pakistan	Independent consultant	November-December 2013
Philippines	Market research firm	April-May 2013
Russian Federation	NGO	September 2013
Thailand	Independent consultant	December 2013
Turkey	Academic research group	September-October 2013
Ukraine	Independent consultant	August 2013
Vietnam	Market research firm	June-July 2013

In-Country Partner Selection

In-country partners were selected on the basis of having good English communication skills (oral and written), a strong background in research and field collection methods, working knowledge of the cities where data collection would take place, and the ability to define a sampling frame for neighborhoods and potential purchase venues in advance of data collection.

Creation of the Sampling Framework for Each Country

The goal in constructing our sampling frame of tobacco vendors in each country was to maximize diversity in packs collected. The country's most populated city and two additional cities of the next nine most populated cities were selected in each country based on cultural, geographic, religious, and linguistic diversity. An underlying philosophy of the Bloomberg Initiative is to focus on places with the greatest number of smokers, so as to maximize the potential impact of any given tobacco control intervention. Given this, we chose to focus data collection on diverse, populous cities within countries of interest. We recognize potential limitations in our sample's focus on populous cities in terms of the possibility of excluding packs of products that are (almost) only consumed in rural areas.

Table 2. Data collection cities per country.

Country	City 1	City 2	City 3
Bangladesh	Dhaka	Sylhet	Chittagong
Brazil	São Paulo	Salvador	Manaus
China[a]	Beijing	Guangzhou	Shanghai
Egypt	Cairo	Alexandria	Mansoura
India	Mumbai	Delhi	Chennai
Indonesia	Jakarta	Semarang	Surabaya
Mexico	Mexico City	Guadalajara	Merida
Pakistan	Islamabad	Lahore	Karachi
Philippines	Manila	Cebu	Davao
Russian Federation	Moscow	Lahore	Karachi
Thailand	Chiang Mai	Bangkok	Hat Yai
Turkey	Istanbul	Diyarbakir	Konya
Ukraine	Kyiv	Lviv	Donetsk
Vietnam	Ho Chi Minh City	Hanoi	Da Nang

[a]In China, data collection was undertaken in 5 cities at the recommendation of tobacco control expert advisors. Additional cities were Kunming and Chengdu.

Training In-Country Field Staff

One of two TPackSS staff traveled traveled to each country to train in-country field staff. Training was delivered in the initial city in which data collection took place in each country. For Egypt and Pakistan, social and political conditions at the time made travel from the United States inadvisable, and, therefore, in-country staff were trained in Dubai, United Arab Emirates. In each country, training took place over five days and included an overview of the project and its goals, hands-on instruction, and supervision with how to carry out data collection procedures, create a data inventory, and take standardized photographs of the tobacco packs collected. TPackSS staff accompanied in-country field staff for purchases in at least four neighborhoods in the initial city. In-country staff were also provided with detailed training reference documents (see Multimedia Appendix/ces 1-4) and remote access to staff for questions throughout the process of data collection, cataloging, and image creation.

Pack Collection

Our aim in this project was to maximize breadth and collect one of every different/unique brand presentation available for sale in the tobacco vendors visited across 36 different neighborhoods in each country.

Within each city, 12 distinct neighborhoods were identified for tobacco vendor sampling. In-country field staff used a variety of local and national resources, including census and property value data, to create a sampling frame of low, moderate, and high socioeconomic areas within the metropolitan boundaries of each city. In each of three socioeconomic strata per city (high, medium, low), we selected four neighborhoods that were diverse in terms of geographic locale and residential composition. Thus, tobacco packs were acquired through purchases made at vendors from a total of 36 different neighborhoods in each country.

Unique packs were defined as tobacco packs that had at least one difference in an exterior feature of the pack. Any pack with a different design or feature, including packs differing in stick count, size, brand name presentation, colors, cellophane, and inclusion of a promotional item was considered to be “unique”. Packs that were exactly the same except for different iterations of the country’s warning labels were not considered to be distinct. Although the majority of packs were easily identified as unique (ie, Marlboro Red was identified as distinct from Marlboro Blue), some packs had minor differences that were more difficult to discern, such as differences in cellophane wrapping around the pack or variations in smaller pack features like brand logo.

In addition to cigarette packs, we also collected cigarros de palha (straw cigarettes from Brazil) and packs sold with promotional items, where these were sold alongside cigarettes. Promotional items were defined to be products that contained a tobacco pack and an additional item such as a lighter, ashtray, or hard tobacco pack carrying case. In countries where bidis or kreteks were sold, these were also purchased.

The initial (index) purchase was always made from a large tobacco vendor in the first sample city, where a broad array of tobacco products was available. We expected the index store purchase to be the largest purchase in each country. In-country staff identified the initial/index vendor in advance of the first day of data collection.

At the initial purchase, every distinct pack available for purchase was acquired, and the price was recorded for each item purchased. The purchase of packs required two field staff. One field staff member worked systematically to review and select every unique cigarette pack on display. The other staff member

recorded price and organized purchased packs so that information on pricing of each pack was retained. After identifying and purchasing one of each distinct pack that was visible, field staff asked the vendor whether there were any additional packs for sale that were not visible, and if available, these packs were also purchased. Where itemized receipts were not provided, staff hand-recorded pack identification numbers (IDs) and price paid before leaving the place of purchase/store. Where permissible, staff took a photo of the tobacco pack display from which packs were identified and purchased. After the purchase, staff used a tablet-based data entry system using the application doForms [24] to record descriptive details about the purchase including city, socioeconomic neighborhood, number of packs purchased, type of vendor, and date of purchase (see Multimedia Appendix 2). Forms were completed and saved and uploaded once Wi-Fi became available.

After the purchase in the initial store, the team returned to the field office where each pack was placed in an individual bag and labeled with a unique ID convention that identified for each pack the country, city, socioeconomic status of the neighborhood, and an assigned pack ID number (see Multimedia Appendix 2). The pack price, type of vendor, and the brand name were also added to the label. Photographs were then taken of the front panel of each pack alongside the ID label and uploaded to a tablet computer to create an archive of purchased packs. Brand families were identified by in-country field staff and were defined as a group of brands (ie, Marlboro Red, Marlboro Blue) that are related and have a parent brand (ie, Marlboro). Brand family folders were created on the tablet, and each purchased pack's front panel image was added to its designated brand family folder. Any pack whose image was in the tablet reference archive was not purchased from subsequent vendors visited. For each country, packs were verified for duplicates upon receipt of the physical packs in Baltimore, Maryland.

The two field staff then visited the 35 remaining neighborhoods (across the 3 cities), and in each neighborhood one vendor was purposively selected based on having a large product inventory. Using the tablet, field staff systematically reviewed all packs available and purchased any pack that did not already appear in the image archive. In the event that the identified vendor did not have any new packs, field staff visited up to three more vendors in the same neighborhood, and made a purchase from the first of these vendors to have any new packs, before moving to the next neighborhood.

Data Inventory Creation

Research Electronic Data Capture (REDCap) [25] was used to manage inventory data. REDCap provides an interface for validated data entry and creates audit trails for tracking data manipulation. In-country field staff undertook initial entry of data on the packs purchased as soon as possible after the purchase was made. IGTC staff checked REDCap data entry regularly during data collection. The data entered on each pack were: pack ID, brand name (Roman as well as any linguistic characters), price, date of purchase, type of tobacco product, manufacturer as presented on the pack, and place of manufacture as presented on the pack. Data access groups were created for each in-country team and user rights were revoked at the completion of data entry to prevent accidental data deletion and to reduce data errors (see Multimedia Appendix 3).

Photographing the Packs

In-country staff created pack images by using a detailed protocol (see Multimedia Appendix 4). Field staff were provided with all photography equipment, including camera, tripod, and lighting equipment. Training on the photography process was covered in one day, with daily practice and review by TPackSS staff during the first week of data collection.

In-country staff took nine standard images of each pack: one direct image of the front panel; one 45 degree angle photo of the front and side panel; one each of the back, top, bottom, and both side panels; one of the opened pack; and one of the cigarette stick. In addition to these nine standard images, in-country staff also captured any other text, including branding on the cellophane and branding revealed when a pack is opened (eg, under the lid). The photos were organized into individual folders by their respective unique identifiers and folders were uploaded to a cloud-based storage application, where they were downloaded by TPackSS staff in Baltimore and checked for image quality. Any images that did not meet protocol standards (eg, improper lighting, unclear text) were identified by TPackSS staff and were retaken by in-country field staff. After initial quality control, TPackSS staff in Baltimore edited each image individually for sizing and brightness necessary for upload to the TPackSS website.

Shipping and Receiving the Packs

All packs were shipped to Baltimore from the country of purchase to facilitate coding for warning label compliance, design features, and pack appeals. Approval was obtained from the US Food and Drug Administration and the Federal Trade Commission to import and store packs for research purposes. Packs were (and continue to be) stored in the TPackSS offices in labeled and catalogued boxes in locked filing cabinets.

As shipments of packs arrived at the TPackSS office in Baltimore, information on the packs was added to a spreadsheet on Google Drive that included the pack unique ID, information about arrival and storage, price paid for the pack, date purchased, vendor of purchase, brand name, brand name in Roman characters (where the initial name was not in Roman characters), tax stamp presence, product type, health warning rotation, verification that the physical pack matched the website pictures, and faces of the pack that had text where translation was necessary. The original inventory data captured by in-country field staff was verified through this process. Missing and duplicate packs were also identified. All packs within a country were systematically checked for the presence of duplicates by comparing each pack to every pack in its respective country.

In the few instances of missing packs upon arrival in Baltimore, staff followed up with in-country partner agencies, and if possible, duplicates of packs missing from the physical dataset were purchased and sent to TPackSS researchers. There were a total of 17 missing packs from the collection upon receipt of shipment (10 from India, 3 from Russia, 1 from Turkey, and 3 from the Philippines). We received 12 replacement packs in

total (8 from India, 3 from the Philippines, and 1 from Turkey), with 5 packs remaining missing (2 from India and 3 from Russia). Of the 5 remaining missing packs, US Customs held 2 because they were manufactured in Cuba.

Using Google Drive allowed for data entry by multiple people simultaneously, and saving a copy of the spreadsheet at the end of each day new data were entered ensured quality. Packs were sorted and stored alphanumerically and by presence of specific health warnings; all packs with a given health warning were stored together for the ease of coding health warning compliance. Packs were resorted by brand name once health warning compliance coding was completed.

Translation

We employed a professional translation service to translate non-English text on the packs. In most cases, it was not necessary to translate warning label text because warning label text on packs could be directly compared with copies of approved warning labels from each country. The translation service was provided with the number of the image as found on the TPackSS website for each pack panel where any translation was required. All translation was entered into a database with rows for each pack and columns ordered by panel. The translation service provided transcription of the text, literal translation, and adaptive translation for cultural meaning when applicable. The translation database was used during coding for compliance with HWL laws and features and appeals (see below).

Creating Codebooks for Pack Compliance With Health Warning Labels Laws

We created a codebook for each country based on the tobacco packaging and labeling laws in effect in each country at the time of data collection. All laws were acquired from the Tobacco Control Laws website [26], a public resource maintained by the Campaign for Tobacco-Free Kids. For countries whose laws were not written in English, we utilized the unofficial English translation of the law, provided on the Tobacco Control Laws website. Only those requirements concerning labeling and packaging were incorporated into the codebooks. Where applicable, the codebooks included measures of inclusion of information on HWLs, emissions and content levels, indications of less harm such as misleading descriptors, and messages prohibiting the sale of cigarettes to minors (see Multimedia Appendix 5).

Any aspect of a law that was considered too nonspecific to be coded consistently was excluded from the codebook. Legal and country experts were consulted when interpretation of the law was in question. Each codebook went through multiple iterations to improve the validity and the reliability of the coding process. During the codebook development process, two members of the research team coded packs in order to assess the reliability of the variables at each stage. Where differences in coding interpretation were identified, codebook questions were clarified. The research team also considered the codebook variables in light of available packs from each country to anticipate coding challenges not obvious from laws alone. Challenges were recorded, discussed, and resolved by the research team.

Creating a Codebook for Features and Appeals

In addition to compliance with tobacco packaging and labeling laws, we also coded each pack for its physical, textual, and visual aspects (features and appeals). Unlike country-specific codebooks to assess compliance with HWL laws, there is one common “Features and Appeals” (F&A) codebook used for all countries (see Multimedia Appendix 6). In order to develop the codebook for F&A, we reviewed the tobacco control literature on packaging and marketing [6,7,27-29]. We also consulted existing coding systems for tobacco packaging F&A, such as the Chatterbox website [30]. We sought to integrate relevant concepts from the published literature on brand appeal, market development, and audience segmentation into our F&A codebook.

Each pack was coded for “features”, which pertain to design elements of the pack including the shape and size of the pack, color, size descriptors, Web presence, the type of opening, and any wrapping or container. The outside and the inside of the pack were both considered, as was the product (stick) itself. We also coded for any evidence of various common “appeals” associated with the tobacco products. Product “appeals” are connections and connotations created in marketing efforts in order to create reasons for a person to purchase (or desire to purchase) a given item (Figure 1 shows this). The TPackSS “appeals” codes are assessments of sociocultural connotations made via various visual elements of branding on the pack to create positive sentiments about the product among a target audience [30]. Product “appeals” can appear both on and inside the pack, on any wrapping or additional packaging, as well as on the stick. We looked for both lexical (words) and images that convey specific appeals. Our appeals codes included (but were not limited to): technology, luxury, femininity, masculinity, youth, nationalism, and United States.

An initial draft of the codebook was developed and then refined to improve objectivity and reliability of coding categories. The coding development and refinement involved test coding sample packs from all 14 countries and a wide variety of pack shapes and opening styles. In addition, we consulted in-country professionals with expertise in tobacco control, communications, and marketing, on the interpretation of culturally significant imagery, and how to objectively code for elements of nationalism and cultural appeal.

Figure 1. Illustrative photo of collection packs.

Coding the Packs Overview

Packs were first coded for compliance with HWL requirements. Only packs displaying a HWL that had been issued by the country in which the pack was purchased and that was in rotation at the time that packs were collected were coded for HWL compliance.

A subsequent and separate process was undertaken for coding packs for F&A. All packs collected (regardless of existence of appropriate HWL) were coded for F&A. In each instance, two independent coders coded every pack. The coders all went through extensive training in tobacco control policy and packaging features, and to the extent possible, the coders were involved in the development of the various codebooks.

TPackSS staff completed all coding using the physical packs rather than the images. To retain a high level of coding quality, coding was limited to approximately 4 hours per day. At the beginning of coding for any country, a team meeting was held to present and discuss the HWL compliance codebook with coders. When necessary, codebooks were revised for clarity and coding issues were resolved.

Over time, one pair of coders specialized in HWL compliance coding, and two sets of coders specialized in F&A coding. All differences or discrepancies within the coding pairs were discussed in regularly held review meetings so as to resolve differences and refine and improve the coding processes. All coding discrepancies and subsequent resolutions were recorded and stored in a central repository. REDCap was used for data entry [25].

In addition to individual pack coding, we also undertook summative and comparative consideration of groups of packs. For HWL compliance, we compared all packs to be coded within each country by warning label and looked for differences in size, color, warning image distortion (such as being stretched or only showing part of the image), placement on the pack, and initial consideration of content. For F&A, we gathered brands across countries and considered all elements of brand consistency and the impact of health warning placement on brand display. In each instance, this was undertaken as a group process with notes taken about notable elements and possible patterns throughout the sets of packs being considered.

Data Management and Quality Control

Database management began with raw data imported from doForms and REDCap and ended with a frozen analytic file with a corresponding codebook. All data were stored in Stata 13 in 2013 to 2014 [31] and then Stata 14 [32] in 2015. Data were accessible only to the database administrator and the database administrator was accountable for all corrections, addition, deletions, and merging of data. Data were routinely backed up on an encrypted external hard drive.

Data audit trails were produced for every addition, deletion, and change of the original data. Multiple data validation checks (eg, extraneous values, outliers, and abnormalities), data verification within a dataset (eg, simple range and constraint validation), and cross-dataset checks (eg, data redundancy and consistency checks) were performed.

For each country, there are five initial datasets that were merged after data cleaning to form a master relational dataset. In order of their generation, the datasets are: (1) Field data; (2) Intake data; (3) HWL compliance; (4) Brand Names and Owners; and (5) F&A. Each pack's unique identifier enabled the merging of different datasets into one master relational database. Field data were verified and cross checked against Intake data. Intake data were verified and cross checked against HWL compliance coding, and so forth. Consulting tobacco brand and tobacco brand owner websites validated brand names. Consulting Euromonitor International [33], a market intelligence report based upon tobacco market research, also validated brand names. For the purpose of this study, "Owner of the Brand" is defined as the entity that holds an active trademark registration of the brand in the particular country and/or would be responsible for the pack in instances of any legal or business related challenges, absent other information. A consultant who utilized portfolios of brands and trademark registry databases validated brand owners.

Because the F&A and HWL compliance data were both entered by independent coding pairs, any discrepancies were reconciled through a process of review by a third trained reviewer. All reconciliations were performed within 24 hours of entering the data for F&A coding and every other day for HWL compliance coding.

Pack measurements (ie, pack height and warning label height) in the HWL compliance codebook required special attention.

Coders measured packs with standardized rules and rounded measurements to the closest millimeter (mm). Given that it was infrequent, although not rare, to get exactly the same measurement between coders, a difference of 1 mm was averaged, while a difference greater than 1 mm required measurement by a third reviewer. Bland-Altman plots were created to assess random error and systematic error in the coder's measurements [34]. Quality checks were also performed on the HWL compliance data to ensure impossible data were not being entered (eg, width of the warning label greater than the width of the pack).

To assess data reliability in the F&A and HWL compliance codebook, Cohen's kappa, prevalence adjusted kappa (PABAK), and interclass correlation coefficient were calculated as appropriate for each country [35-39]. Each coder's entry for a record was compared against the final merged record to assess how frequently coders were agreeing with the final record. The PABAK statistic takes into account low-prevalence bias that skews the Cohen's kappa. By evaluating both of these statistics, we identified variables with low agreement (< 80% agreement). A 2-sided alpha of < .05 was considered statistically significant. Variables with low agreement were flagged and coders were routinely provided feedback on variables for which there was low agreement. This feedback was used to clarify instructions for the codebook. Follow-up was performed to ensure agreement of these variables improved as coding progressed.

Developing a Health Warning Label Compliance Score

Article 11 of the FCTC outlines mechanisms by which parties to the treaty can increase the effectiveness of their tobacco packaging and labeling. Key elements include location; size; use of pictorials; color; rotation; message content; language; source attribution; and information on constituents and emissions.

We operationalized HWL compliance through four overarching categories that related to requirements across the study countries: (1) Warning location; (2) Warning size; (3) Text size in the warning; and (4) Warning label elements (such as color or content of warning). All four key requirements were assessed for 10 countries. For the four remaining countries, only 3 of the 4 categories pertained to the country's law, and so only these elements were included in the compliance measure (Table 3). The compliance score was calculated as number of compliant packs divided by number of packs with a warning label in rotation at the time of collection. In addition, a composite compliance score was determined by dividing the number of packs compliant on all four key requirements by the total number of packs coded for each country. The compliance results are outside of the scope of this protocol paper.

Table 3. HWL compliance measures as related to elements of countries' laws.

Country	Warning type & location	Warning size	Warning text size	Warning label elements
Bangladesh	Text warnings on front and back of the pack	30% of the front and back	18 point font	Black text on white background or white text on black background
Brazil	Picture-based warning on back of the pack	100% of the back	Proportion and graphic parameters of images provided by the Brazilian Health Surveillance Agency must be unchanged	White text on black rectangular background
China	Text warnings on front and back of the pack	30% of the front and back	4 mm tall text	Color contrast between the text and background
Egypt	Picture-based warning on front and back of the pack	50% of the front and back	Not applicable	Text printed on black background; quit line and standard warning printed on yellow background
India	Picture-based warning on front of the pack	40% of the front	Warning must be 0.75 to 1 ratio of its vertical to horizontal length	Red and white text on black background
Indonesia	Text warnings on a "part of the package that is easily read"	Not applicable	3 mm tall text	Black text on background that is a shade of white with black border
Mexico	Picture-based warning on front and text warning on back of the pack	30% of the front and 100% of the back	10 point font on front; 9-11 point font on back	Yellow text on front and back; black background on back
Pakistan	Picture-based warnings on front and back of the pack	40% of the front and back	2 mm tall text	Black text on white background
Philippines	Text warnings on front of the pack	30% of the front	Text must comprise at least 50% of the warning	Black text on white background with black border
Russian Federation	Text warning on front and picture-based warning on back of the pack	30% of the front and 50% of the back	Not applicable	Black border on front and back; black text on white background on back
Thailand	Picture-based warnings on front and back of the pack	55% of the front and back	Size and positioning of text must appear as it does in examples provided by the Ministry of Health	Content must appear as it does in examples provided by the Ministry of Health
Turkey	Picture-based warning on front and picture-based warning on back of the pack	65% of the front and back	Not applicable	Black border on front and back; black text on back
Ukraine	Text warning on front and picture-based warning on back of the pack	50% of the front and back	Must occupy no less than 40% of the area within the black border of the health warning	Black text on white background with black border
Vietnam	Text warnings on front and back of the pack	30% of the front and back	2 mm tall text	Black text on white background

Results

Packs Collected

Across the 14 countries, we collected 3307 tobacco packs. We collected 3006 cigarette packs, 55 bidi packs, 234 kretek packs, 3 cigarros de palha (straw cigarettes), and 9 promotional items. Product type was determined by labeling on the pack. We collected the most packs from the Russian Federation (n=505) and the fewest packs from Egypt (n=58). We will be analyzing data on compliance with HWL requirements, pack F&A, and pricing of products in separate analyses to be published at a later date.

Table 4. Packs collected in each country and each city.

Country		Number of packs collected		
	Total	City 1	City 2	City 3
Bangladesh	200	143	11	46
Brazil	130	104	10	16
China[a]	453	227	70	55
Egypt	58	58	-	-
India	169	108	29	32
Indonesia	215	115	35	65
Mexico	134	107	19	8
Pakistan	394	296	58	40
Philippines	144	79	53	12
Russian Federation	505	406	61	38
Thailand	126	57	53	16
Turkey	308	206	72	30
Ukraine	324	242	48	34
Vietnam	147	120	24	3
Total	3307[a]	2268	543	395

[a]Total includes packs from China city 4=50 & city 5=51 packs

Creating a Website

The TPackSS searchable Internet archive is now publicly available [40] (Figure 2 shows this). This dynamic archive houses images of all the tobacco packs purchased in the 14 countries. The intention is for this archive to be used by the tobacco control community to monitor compliance with existing HWL laws, understand innovation in pack design and brand promotion, and advocate for policy change that can prevent future harm from tobacco use.

A Web development company was contracted to create a site that can be searched and filtered by country, brand family, brand owner, and tobacco product type. Pack specific data such as brand family, product type, price, purchase date, purchase city, and purchase country were paired with each pack and are also available on the site. The site includes information about packaging and labeling regulations in each country.

Future plans for the website include uploading the key compliance, design, and F&A variables for each pack, and incorporating the capacity to view the site in multiple languages.

Figure 2. Screenshot of Internet archive.

Discussion

TPackSS data collection will be repeated at a minimum of two years following the previous data collection, for those countries where the warning label or packaging requirements have changed. This will facilitate comparisons of packaging pre and post policy implementation.

In addition to creating the Internet archive, we are compiling country-specific fact sheets on HWL compliance to support policy advocacy. We are also creating resources (country specific and comparative) related to various aspects of the F&A. The fact sheets will be downloadable from the TPackSS website [40]. We have made our field protocol documents, health warning compliance codebooks, and F&A codebook available on the project website.

Acknowledgments

This study would not be possible without the entire TPackSS team, including our coders Ashley Conroy, Cerise Kleb, and Adamove Osho. We would also like to thank all of our in-country field partners for their assistance in this study, as well as the numerous tobacco control experts from various countries for input on cultural references being made on the tobacco packs. We also acknowledge the work of Dr Hana Ross who provided data on brand ownership. Finally, we acknowledge all of the expert consultants who assisted us in our initial planning of the study.

This work was supported by a grant from the Bloomberg Initiative to Reduce Tobacco Use to the Johns Hopkins Bloomberg School of Public Health.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Training manual.

[PDF File (Adobe PDF File), 1MB - publichealth_v1i2e8_app1.pdf]

Multimedia Appendix 2

In-field data collection form for vendor and neighborhood.

[PDF File (Adobe PDF File), 329KB - publichealth_v1i2e8_app2.pdf]

Multimedia Appendix 3

REDCap in-country inventory form.

[PDF File (Adobe PDF File), 865KB - publichealth_v1i2e8_app3.pdf]

Multimedia Appendix 4

Photography protocol.

[PDF File (Adobe PDF File), 1MB - publichealth_v1i2e8_app4.pdf]

Multimedia Appendix 5

Example of one country health warning compliance codebook.

[PDF File (Adobe PDF File), 1MB - publichealth_v1i2e8_app5.pdf]

Multimedia Appendix 6

Features and appeals codebook.

[PDF File (Adobe PDF File), 3MB - publichealth_v1i2e8_app6.pdf]

References

1. World Health Organization. WHO report on the global tobacco epidemic, 2013: enforcing bans on tobacco advertising, promotion and sponsorship URL: http://apps.who.int/iris/bitstream/10665/85380/1/9789241505871_eng.pdf?ua=1 [accessed 2015-08-04] [WebCite Cache ID 6aXRufYiV]
2. Wipfli H. American Journal of Epidemiology. Atlanta, GA: American Cancer Society; New York, NY: World Lung Foundation; 2012 Nov 06. The tobacco atlas, fourth edition URL: http://aje.oxfordjournals.org/content/early/2012/11/05/aje.kws389.full [accessed 2015-08-04] [WebCite Cache ID 6aXSPYdzO]
3. World Health Organization. The WHO framework convention on tobacco control: An overview URL: http://www.who.int/fctc/about/WHO_FCTC_summary_January2015.pdf [accessed 2015-04-15] [WebCite Cache ID 6XoBW3iqb]
4. Wakefield M, Germain D, Durkin S, Hammond D, Goldberg M, Borland R. Do larger pictorial health warnings diminish the need for plain packaging of cigarettes? Addiction 2012 Jun;107(6):1159-1167. [doi: 10.1111/j.1360-0443.2012.03774.x] [Medline: 22372966]
5. Hammond D. Health warning messages on tobacco products: A review. Tob Control 2011 Sep;20(5):327-337. [doi: 10.1136/tc.2010.037630] [Medline: 21606180]
6. Hammond D, Dockrell M, Arnott D, Lee A, McNeill A. Cigarette pack design and perceptions of risk among UK adults and youth. Eur J Public Health 2009 Dec;19(6):631-637 [FREE Full text] [doi: 10.1093/eurpub/ckp122] [Medline: 19726589]
7. Wakefield M, Morley C, Horan JK, Cummings KM. The cigarette pack as image: New evidence from tobacco industry documents. Tob Control 2002 Mar;11 Suppl 1:I73-I80 [FREE Full text] [Medline: 11893817]
8. Institute for Global Tobacco Control. State of evidence review: Health warning labels on tobacco products URL: http://globaltobaccocontrol.org/sites/default/files/HealthWarnings_state_of_evidence_final_11_18_2013_web_0.pdf [accessed 2015-04-15] [WebCite Cache ID 6XoSFe0a1]
9. Bansal-Travers M, Hammond D, Smith P, Cummings KM. The impact of cigarette pack design, descriptors, and warning labels on risk perception in the U.S. Am J Prev Med 2011 Jun;40(6):674-682 [FREE Full text] [doi: 10.1016/j.amepre.2011.01.021] [Medline: 21565661]
10. Cancer society. Canadian: Cancer society, Cigarette package health warnings: International status report, fourth edition, September 2014 URL: http://www.tobaccolabels.ca/wp/wp-content/uploads/2014/10/Cigarette-Package-Health-Warnings-International-Status-Report-English-CCS-Sept-2014.pdf [accessed 2015-04-15] [WebCite Cache ID 6XoApX6RD]
11. Partos TR, Borland R, Yong H, Thrasher J, Hammond D. Cigarette packet warning labels can prevent relapse: Findings from the International Tobacco Control 4-Country policy evaluation cohort study. Tob Control 2013 May;22(e1):e43-e50 [FREE Full text] [doi: 10.1136/tobaccocontrol-2011-050254] [Medline: 22535363]
12. Hammond D, Fong GT, McNeill A, Borland R, Cummings KM. Effectiveness of cigarette warning labels in informing smokers about the risks of smoking: Findings from the International Tobacco Control (ITC) Four Country Survey. Tob Control 2006 Jun;15 Suppl 3:iii19-iii25 [FREE Full text] [doi: 10.1136/tc.2005.012294] [Medline: 16754942]
13. Fathelrahman AI, Omar M, Awang R, Borland R, Fong GT, Hammond D, et al. Smokers' responses toward cigarette pack warning labels in predicting quit intention, stage of change, and self-efficacy. Nicotine Tob Res 2009 Mar;11(3):248-253 [FREE Full text] [doi: 10.1093/ntr/ntn029] [Medline: 19246625]

14. Borland R, Yong HH, Wilson N, Fong GT, Hammond D, Cummings KM, et al. How reactions to cigarette packet health warnings influence quitting: Findings from the ITC Four-Country survey. Addiction 2009 Apr;104(4):669-675 [FREE Full text] [doi: 10.1111/j.1360-0443.2009.02508.x] [Medline: 19215595]
15. Hammond D, Wakefield M, Durkin S, Brennan E. Tobacco packaging and mass media campaigns: Research needs for Articles 11 and 12 of the WHO Framework Convention on Tobacco Control. Nicotine Tob Res 2013 Apr;15(4):817-831 [FREE Full text] [doi: 10.1093/ntr/nts202] [Medline: 23042986]
16. Fong GT, Hammond D, Hitchman SC. The impact of pictures on the effectiveness of tobacco warnings. Bull World Health Organ 2009 Aug;87(8):640-643 [FREE Full text] [Medline: 19705020]
17. Borland R, Wilson N, Fong GT, Hammond D, Cummings KM, Yong H, et al. Impact of graphic and text warnings on cigarette packs: Findings from four countries over five years. Tob Control 2009 Oct;18(5):358-364. [doi: 10.1136/tc.2008.028043] [Medline: 19561362]
18. Hammond D, Fong GT, Borland R, Cummings KM, McNeill A, Driezen P. Text and graphic warnings on cigarette packages: Findings from the international tobacco control four country study. Am J Prev Med 2007 Mar;32(3):202-209 [FREE Full text] [doi: 10.1016/j.amepre.2006.11.011] [Medline: 17296472]
19. Wang A, Lowen SB, Romer D, Giorno M, Langleben DD. Emotional reaction facilitates the brain and behavioural impact of graphic cigarette warning labels in smokers. Tob Control 2015 May;24(3):225-232. [doi: 10.1136/tobaccocontrol-2014-051993] [Medline: 25564288]
20. Shiffman S, Pillitteri JL, Burton SL, Rohay JM, Gitchell JG. Smokers' beliefs about "Light" and "Ultra Light" cigarettes. Tob Control 2001;10 Suppl 1:i17-i23 [FREE Full text] [Medline: 11740040]
21. Hamilton WL, Norton GD, Ouellette TK, Rhodes WM, Kling R, Connolly GN. Smokers' responses to advertisements for regular and light cigarettes and potential reduced-exposure tobacco products. Nicotine Tob Res 2004 Dec;6 Suppl 3:S353-S362. [Medline: 15799598]
22. Cohen JE, Yang J, Donaldson EA. Impact of the removal of light and mild descriptors from cigarette packages in Ontario, Canada: Switching to "light replacement" brand variants. Prev Med 2014 Dec;69:120-125 [FREE Full text] [doi: 10.1016/j.ypmed.2014.08.037] [Medline: 25224153]
23. World HO. Bloomberg initiative components. Tobacco free initiative URL: http://www.who.int/tobacco/about/partners/bloomberg/initiative_components/en/ [accessed 2015-04-15] [WebCite Cache ID 6XoRyZ9XX]
24. doForms. 2015-04-15 URL: http://www.doforms.com/product-overview [accessed 2015-04-15] [WebCite Cache ID 6XoS9Vtmv]
25. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap)--a metadata-driven methodology and workflow process for providing translational research informatics support. J Biomed Inform 2009 Apr;42(2):377-381 [FREE Full text] [doi: 10.1016/j.jbi.2008.08.010] [Medline: 18929686]
26. Campaign for Tobacco Free Kids. Tobacco control laws URL: http://www.tobaccocontrollaws.org/ [accessed 2015-04-15] [WebCite Cache ID 6XoHMSLr8]
27. Ford A, Moodie C, Hastings G. The role of packaging for consumer products: Understanding the move towards ‘plain’ tobacco packaging. Addict Res Theory 2012 Aug;20(4):339-347. [doi: 10.3109/16066359.2011.632700]
28. Hammond D, Doxey J, Daniel S, Bansal-Travers M. Impact of female-oriented cigarette packaging in the United States. Nicotine Tob Res 2011 Jul;13(7):579-588 [FREE Full text] [doi: 10.1093/ntr/ntr045] [Medline: 21486994]
29. Dewhirst T, Lee WB. Cigarette advertising in the Republic of Korea: A case illustration of The One. Tob Control 2012 Nov;21(6):584-588. [doi: 10.1136/tobaccocontrol-2011-050315] [Medline: 22517803]
30. Health Canada. Chatter Box. URL: http://chatterbox.otru.org/ [accessed 2015-04-15] [WebCite Cache ID 6XoNEpcTK]
31. StataCorp. StataCorp. College Station, TX: Release 13. : StataCorp LP; 2013.
32. StataCorp. StataCorp. College Station, TX: Release 14. : StataCorp LP; 2015. URL: http://www.stata.com/manuals14/u.pdf [accessed 2015-08-04] [WebCite Cache ID 6aXWIuvUm]
33. Euromonitor. Retrieved from Euromonitor Passport Database. 2013. Tobacco statistics by country URL: http://www.euromonitor.com/tobacco [WebCite Cache ID 6aXXH7xWB]
34. Altman D, Bland J. Statistician. Measurement in medicine: the analysis of method comparison studies. Statistician ; 1983:307-317.
35. Byrt T, Bishop J, Carlin JB. Bias, prevalence and kappa. J Clin Epidemiol 1993 May;46(5):423-429. [Medline: 8501467]
36. Gwet KL. Computing inter-rater reliability and its variance in the presence of high agreement. Br J Math Stat Psychol 2008 May;61(Pt 1):29-48. [doi: 10.1348/000711006X126600] [Medline: 18482474]
37. Viera AJ, Garrett JM. Understanding interobserver agreement: The kappa statistic. Fam Med 2005 May;37(5):360-363 [FREE Full text] [Medline: 15883903]
38. Hallgren KA. Computing inter-rater reliability for observational data: An overview and tutorial. Tutor Quant Methods Psychol 2012;8(1):23-34 [FREE Full text] [Medline: 22833776]
39. Cunningham M. SAS global forum. More than just the kappa coefficient: A program to fully characterize inter-rater reliability between two raters URL: http://support.sas.com/resources/papers/proceedings09/242-2009.pdf [accessed 2015-04-15] [WebCite Cache ID 6XoRrA7qH]

40. Institute for Global Tobacco Control. Tobacco pack surveillance system:TPackSS URL: http://globaltobaccocontrol.org/tpackss/ [accessed 2015-04-15] [WebCite Cache ID 6XoNUkHUm]

Abbreviations

F&A: features and appeals
FCTC: Framework Convention on Tobacco Control
IGTC: Global Tobacco Control
HWLs: health warning labels
IDs: identification numbers
mm: millimeter
NGO: nongovernmental organization
PABAK: prevalence adjusted kappa
REDCap: Research Electronic Data Capture
TPackSS: Tobacco Pack Surveillance System

Edited by G Eysenbach; submitted 13.05.15; peer-reviewed by J Kimmerle, MP Wang; comments to author 19.06.15; revised version received 02.07.15; accepted 03.07.15; published 12.08.15.

Please cite as:
Smith K, Washington C, Brown J, Vadnais A, Kroart L, Ferguson J, Cohen J
The Tobacco Pack Surveillance System: A Protocol for Assessing Health Warning Compliance, Design Features, and Appeals of Tobacco Packs Sold in Low- and Middle-Income Countries
JMIR Public Health Surveill 2015;1(2):e8
URL: http://publichealth.jmir.org/2015/2/e8/
doi:10.2196/publichealth.4616
PMID:27227142

Original Paper

A Qualitative Examination of Respondent-Driven Sampling (RDS) Peer Referral Challenges Among Young Transwomen in the San Francisco Bay Area

Sean Arayasirikul[1,2]; Xiang Cai[3], BA; Erin C Wilson[1,4], DrPH

[1]Center for Public Health, Population Health Division, San Francisco Department of Public Health, San Francisco, CA, United States

[2]Medical Sociology, Department of Social and Behavioral Sciences, University of California San Francisco, San Francisco, CA, United States

[3]Yale University, New Haven, CT, United States

[4]Department of Epidemiology and Biostatistics, University of California San Francisco, San Francisco, CA, United States

Corresponding Author:
Sean Arayasirikul
Center for Public Health
Population Health Division
San Francisco Department of Public Health
25 Van Ness Ave 5th Floor
San Francisco, CA, 94102
United States
Phone: 1 4155549036
Fax: 1 4155549000
Email: sean.arayasirikul@gmail.com

Abstract

Background: Efforts have focused on developing innovative recruitment strategies to engage the most marginalized of populations in public health research. Respondent-driven sampling (RDS) has been found to be an effective sampling strategy for hard-to-reach, hidden populations. Though studies have documented RDS peer referral as challenging, literature contextualizing these challenges is scant and rarely do they discuss the role of Internet technologies.

Objective: The objective of the study was to explore reasons for peer referral challenges in a human immunodeficiency virus (HIV) risk and resilience study among a hidden population of youth, specifically, young transwomen. These findings amplify the unique opportunities Internet technologies bring to public health research and methodology.

Methods: We conducted focused, semistructured, qualitative interviews with 16 young transwomen to investigate the reasons why youth did or did not refer peers to an RDS study for transwomen ages 16-24 in the San Francisco Bay Area. Qualitative interview data were coded and analyzed using grounded theory.

Results: Participants discussed specific barriers and facilitators related to four factors that include study design, study implementation, community characteristics, and individual characteristics, which contributed to RDS peer referral challenges.

Conclusions: Our grounded theory analysis identifies important considerations for future RDS studies with hidden youth populations. Exploring research participants' experiences is integral in strengthening future epidemiologic research efforts that plan to use RDS to sample and estimate the hidden epidemics among at-risk youth and transgender women. Additionally, Internet technologies and Web-based adaptations offer solutions to traditional RDS peer referral challenges, having the potential to increase accessibility and use among hidden youth populations.

(JMIR Public Health Surveill 2015;1(2):e9) doi:10.2196/publichealth.4573

KEYWORDS

transgenders; young adults; qualitative research; epidemiology; methodology

Introduction

Accessing Transgender Women

Access to hard-to-reach, hidden populations is often limited, but necessary to characterize epidemics among key populations at risk for human immunodeficiency virus (HIV) [1,2]. However, such populations are often difficult to sample [3]. Respondent-driven sampling (RDS), an adaptation of chain-referral sampling [4], has been used to sample diverse hard-to-reach, hidden populations. Transgender women are a population disproportionately impacted by HIV around the world [5] and are considered a hidden population due to gender-based stigma toward this group. RDS studies have been quite successful in reaching transwomen, individuals who not identify with the gender associated with their assigned male sex at birth, for HIV research [6-9], but gaps in specific subpopulations remain.

Research has emerged finding that younger transwomen are also at high risk for HIV [10], but there have been calls for more rigorous population-based studies to assess local epidemics and determine current risk. Younger people in general are a group particularly difficult to recruit using RDS [11-15]. A RDS-evaluation study comparing field data to population data found that being younger was one of only four factors associated with being underrepresented in the RDS data [15]. RDS for finding hidden youth populations has been even more problematic. For example, it took 12 waves of recruitment to achieve a sample size of 259 young women with multiple sex partners in a South African study [12]. It took over three years to recruit a sample of 450 young men who have sex with men (YMSM) using respondent driven sampling in a recent U.S. study [13]. The only factor significantly related to recruitment success in the YMSM study was having a large network size. Some research has been done to identify ways to overcome these issues and to determine what demographic factors are associated, which may give some insight into what populations may be most challenging [13]. No studies were found in the literature that qualitatively investigated reasons for challenges associated with RDS referral among a hidden youth population.

Purpose of the Analysis

The purpose of this analysis was to fill a gap in the literature by examining the referral experiences of youth participants in an RDS study. Specifically, we sought to address the following research question, “What barriers and facilitators did young transwomen encounter in RDS peer referral?” We discuss how these findings amplify the unique opportunities Internet technologies bring to public health research and methodology and considerations for future applications of RDS peer referral among hidden youth populations.

Methods

The SHINE Study

The SHINE study is the first longitudinal study, to our knowledge, of HIV risk and resilience among young transwomen, ages 16-24 [16], in the San Francisco Bay Area. We conducted six focus groups with participants and confirmed feasibility and acceptability of RDS in this hidden, hard-to-reach population. After nine months of RDS implementation, seeds were not propagating, leading to few peer referrals. In order to boost recruitment, we later incorporated direct referrals from community-based organizations, outreach at events, and online outreach through social networks to identify new seeds until a cohort of 300 individuals were enrolled. Participants were compensated for their participation at each data collection time point (baseline, 6-month, and 12-month) in the amounts of US $50, $70, and $100, respectively. At the end of participants’ baseline visit, study staff explained RDS peer referral procedures. Participants were provided with three referral coupons and earned US $20 for each successful referral.

Participant Recruitment, Procedures, and Analysis

We conducted focused, semistructured, qualitative interviews with a subsample of 16 participants of the parent study via telephone. Participants were purposively sampled in order to obtain diversity in ability/willingness to provide study referrals to peers in age, race/ethnicity, and socioeconomic status. Participants were not paid. Interviews lasted 10-15 minutes and took place during a time that was most convenient for the participant. The interview guide was iterated in order to maximize coverage of participant experiences through theoretical sampling to reach theoretical saturation [17] and to address the following research question, “What barriers and facilitators did young transwomen encounter in RDS peer referral?” The interview guide assessed the following constructs: friendship networks, social isolation; knowledge, attitudes, motivation, and behavior related to peer referral; and peer referral successes, challenges, and improvements.

Interviews were audio-recorded and transcribed verbatim by the second author. Transcriptions were randomly checked for quality and accuracy against original recordings by the first author. Qualitative interview data were coded and analyzed using grounded theory [17]. The first and second authors independently coded qualitative data, line by line, and together, organized codes into categories to identify specific factors that influenced RDS referrals among participants. The last author oversaw all research procedures and is an expert in qualitative research. The Institutional Review Board (IRB) at the University of California, San Francisco, approved all study procedures. All participants provided written consent (or written assent for those younger than 18 years of age in accordance with a review board waiver of parental consent) to participate in the longitudinal parent study. We obtained IRB approval six months later to approach consented participants in the longitudinal parent study to recruit for this substudy in which verbal consent was approved for and obtained.

Results

The Participants

The age range of the 16 participants was 17 to 24 years old, with a mean age of 21.25 years. The majority, (10/16) of the subsample, were young transwomen of color, with 12% (2/16) identifying as African American, 25% (4/16) as Asian/Pacific Islander, 19% (3/16) as Latina, and 6% (1/16) as mixed race. There were (50%) 8/16 participants that reported having

completed a high school education or less. There were (50%) 8/16 participants that reported a monthly income of US $0-$500. About a third (n=5) of participants were unable to refer a peer, and two-thirds referred 1 or more peers.

Qualitative data revealed specific barriers and facilitators participants encountered while referring their peers. These factors include study design, study implementation, community characteristics, and individual characteristics to explain participants' perspectives on RDS peer referral. Table 1 is a qualitative matrix showing how participants' interviews (in columns) were categorized across factors and specific codes (in rows). Table 1 is organized by the number of peer referrals participants made, ranging from unsuccessful recruiters (or those who referred no peers), to moderately successful recruiters (or those who referred 1-2 peers), and, finally, successful recruiters (or those who referred 3 or more peers). For example, Participant A was an unsuccessful recruiter, who made no peer referrals, and reported that the incentives for peer referrals were inadequate (denoted by the "x" in that row for that column) in the Study design factor. Participant A went on to report that the referral process was confusing and that the community was small (in the Study implementation and Community characteristics factors, respectively). Table 1 presents the qualitative analysis of factors that served as barriers or facilitators to participants' ability to refer their peers.

Table 1. Qualitative matrix of factors and codes by participant and their number of peer referrals.

	Unsuccessful recruiters					Moderately successful recruiters					Successful recruiters					
Number of referrals[a]	0					1				2	3			5		7
Participant	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
Matrix of factors																
Study design																
Narrow age eligibility			x	x				x				x		x		
Inadequate incentives	x		x													x
Adequate incentives		x		x	x	x	x		x	x	x	x	x		x	
Study implementation																
Paper coupons-difficult to retain					x		x									
Paper coupons not effective					x					x						
Referral process confusing	x	x			x	x		x		x				x		
Maintaining a close relationship with staff														x	x	x
Actively following up with peers						x					x	x			x	x
Community characteristics																
Small community size	x	x	x	x		x	x	x	x	x				x	x	
Sampling saturation			x	x	x	x	x	x	x	x		x				
Participation as stigmatizing						x				x						
Individual characteristics																
Social anxiety and discomfort		x			x	x		x								
Altruism and reciprocity			x						x	x	x		x	x		x

[a]Each column noted by letter represents one participant.

Study Design

The study's eligibility criteria created a challenge for participants who wanted to refer other peers. There were five participants that reported that the study's age eligibility criterion was too narrow (ages 16-24 years old). This negatively impacted participants' ability to refer peers who were too old for the study, but who made up a substantial part of their social network. There was one participant that reported little access to social settings where other young transwomen were present. She explained,

> *I am in the upper limit of the age range for the SHINE study, and most of my friends are [older]. You know, I don't really hang out with younger trans youth.* [Study participant]

Participants reported the study's double incentivizing system as both a facilitator and a barrier for referring peers. The US $20 per enrolled participant referral incentive was seen as a good or adequate amount for 11 participants. Though, three participants expressed that the incentive was inadequate. There was one participant that explained that in relation to the amount of work and time it might take to make a successful referral, the incentive was less motivating.

Study Implementation

Participants identified the study's referral coupons, referral training, and the relationship with study staff as important to their ability to recruit their peers. For example, three participants who had referred 5 or more peers noted that maintaining a close relationship with research staff supported their referral efforts. There was one participant that even shared that the rapport she developed with research staff over time motivated her to actively seek out potential participants in their network. This participant said that as a result of rapport with study staff, she sought "to be the best possible referral person ever". Additionally, good rapport with participants helped participants form a better understanding of the study, its objectives, and eligibility criteria. There was one participant that said,

> *I definitely felt more comfortable when I was completely brought on board with what was going on and what was it about as well as what to say and how to get people in.* [Study participant]

Additionally, five participants who successfully referred peers were more likely to report practicing active follow-up with peers they gave coupons to. Types of active follow-up these participants implemented included telephone calls, short message service texts, and accompanying referrals to their study visits. Some took the initiative to schedule appointments with study staff for their peers to ensure their follow through.

For two participants, paper coupons were difficult to retain because they were struggling with residential instability; thus, holding onto paper coupons was challenging. There were two participants that reported that the paper coupon was not an effective referral tool. There was one participant that said,

> *People disregard most of the coupons nowadays. Just like okay, I will just put it somewhere and forget about it.* [Study participant]

Additionally, seven participants reported that they were confused about how to refer peers or how important it was to refer; as a result, participants often felt little responsibility to refer peers. There was one participant that said,

> *I am not going to go to somebody and like 'Hey, how is it going?' and just hand them out coupons. It is just a little weird, I think.* [Study participant]

Another participant said that she "was not particularly sure how much information [she] could give out". Whereas another participant said,

> *It wasn't in mind that I needed to tell them. I feel like I was not so informed about the SHINE Study, and I didn't think I needed to recruit.* [Study participant]

Community Characteristics

Participants identified a number of important community/population-specific factors that impacted their ability to refer peers. Specifically, the size of the community, sampling saturation, and participation in research as stigmatizing, emerged as barriers to successful referrals.

There were 11 participants that reported that the young transwomen population overall was very small, which made it very difficult to refer peers. For those who lived outside the metropolitan areas of San Francisco and Oakland, referrals were particularly difficult because the population was even smaller. There was one participant that explained,

> *I think a lot of girls who also participated in this study live in different areas of the Bay Area. So if you don't live in San Francisco or you don't live in Oakland, your network of girls would be a lot smaller.* [Study participant]

There were 9 participants that reported that the study had reached a point of sampling saturation, saying that most or all of their friends were already enrolled. There was one participant that commented, "I know a lot of people, but a lot of the people I reached out to have already contacted the SHINE study". Another participant shared, "the challenge [with referring peers] is finding someone who hasn't already done the SHINE Study".

Additionally, two participants expressed the belief that participation in research can be stigmatizing. There was one participant that elaborated,

> *I think [young transwomen] are vulnerable [because] like the study is like a medical experience, like a mental health experience. I think a lot of trans people have a negative impression toward the medical profession. Or like just a general anxiety about it. Like I don't know, it might feel like going to the doctor without really meaning to, I guess.* [Study participant]

These participants were influenced by negative experiences with medical and mental health institutions in the past and a deep concern around protection of one's privacy.

Individual Characteristics

At the individual level, a number of participants reported experiencing social discomfort around RDS referrals, while others reported altruism as a motivator for peer referral. There were four participants that expressed that they had feelings of social anxiety and discomfort associated with referring peers. There was one participant that explained,

> *People stress me out, and I don't like to talk to people. I don't get social cues. I don't know; I am just bad at it. I just like to go home and watch Netflix. It's just not my personality to try to get people to do things. That is why I am not in sales. That is why I sit in front of a computer and program all day, where I don't have to talk to people.* [Study participant]

Participants also reported altruism and reciprocity, which helped to motivate their own participation in the study, and subsequently helped motivate some youth to make peer referrals. There was one participant that said,

> *[The study] is pretty fascinating. In like since it is specifically for [young] transwomen, I was pretty impressed. So like I think, it was an honor to be a part of it...I would be willing to do it without receiving anything. I don't really care for it. Because, I like, I feel like the purpose and the goal is more important.* [Study participant]

There were seven participants that reported that their own participation in this study was motivated by helping the community of young transwomen and they, in turn, would benefit from the impact of the study's findings.

Discussion

Principal Findings

These data identify specific factors that serve as barriers and facilitators to RDS peer referral among young transwomen and suggest important considerations for future RDS studies with hidden youth populations. Contrary to adult RDS studies that found low monetary incentives to be inadequate for generating peer referrals, data in this study found that the majority of participants reported the US $20 incentive as adequate. Previous RDS studies conducted in adult populations have experienced challenges in successfully incentivizing peer referrals [18,19]. An RDS study of HIV risk among international MSM travelers offered a secondary incentive for peer referrals in the amount of US $10 [18]. Participants in this study did not respond to that level of monetary incentivization, which prompted study investigators to instead offer participants a raffle entry for prizes of US $500. Another study found that even at that level of incentivization, raffling large monetary sums was not effective among an adult sample of cannabis users [19]. Finding that this level of incentivization was effective for youth supports future RDS studies with youth populations.

With regard to RDS implementation, participants identified challenges around the use of paper coupons and confusion around the referral process. Methodological developments in RDS studies have expanded to include the use of Web adaptations of RDS, referred to as webRDS, which may help to address these challenges [11,20]. WebRDS uses an Internet portal to assign unique identifiers to participants and enable them to generate electronic coupons and linked email messages, which can be sent to peer referrals [20]. Because few public health studies have utilized webRDS [11,18,21-23], a rigorous exploration of webRDS among hidden youth populations may ameliorate challenges we found associated with the use of paper coupons. There was one study that observed that when presented with the choice, study participants strongly preferred electronic coupons to paper coupons [18]. The success of electronic coupons in this study has even suggested that electronic referrals may enhance random selection of peers if participants were not limited to in-person meetings to transfer paper coupons [18].

Broadly, our data found that youth participants were influenced by research-related, relationship, and interactional factors rather than the adequacy of monetary incentives. For example, we found that participants who maintained close working relationships with research staff were more motivated and, as a result, able to successfully refer peers. There was one study that found that though youth often want the potential benefits of research, they vary in their cognitive ability to understand important research details and procedures [24]. Our findings underscore the importance of providing a supportive environment with multiple engagement opportunities for youth participating in research, similar to other published work calling for increased engagement with youth and adolescent communities to address perceived barriers to participation [25,26].

Related to research participation in general, we found that youth participating in research can be stigmatizing and is important for understanding the use of RDS among hidden populations. This is especially true for research studies with vulnerable and marginalized youth that rely on the assumption that participants be "out" as trans, at the very least to themselves and the researchers [27]. Literature has identified many barriers to research participation among socially disadvantaged groups, such as medical mistrust, fear of authority, stigma, mistreatment, or exploitation; these reasons were especially salient for racial, gender, and sexual minorities [3,28-30]. Moreover, HIV research studies have struggled to sample adolescents due to HIV-related perceived stigma and negative social consequences [27]. WebRDS may create opportunities to address these larger research-related issues, making research more accessible and youth-friendly [11]. WebRDS has been found to address some of these challenges around communicating with potential peer referrals by affording youth the ability to recruit peers through passive or active strategies, using the approach that they prefer most [11]. WebRDS may alleviate the social anxiety and discomfort participants reported related to recruiting peers and possibly aid in protecting their anonymity or the confidentiality of their gender identity [11].

Limitations

There are a number of limitations to this study. The small size of this subsample limits the generalizability of these results to both the large cohort of the parent study as well as the population. Though purposive sampling was used to generate a range of experiences in recruitment, it is subject to selection bias. The extent to which social desirability bias influenced participants' discussion with researchers about the quality of the research experience is a possibility. Additionally, these data are constrained by brief one-time interviews centered on a specific topic, RDS peer referral. Despite these limitations, these findings highlight the import in examining the experiences of research participants themselves. Most importantly, these data seek to inform and bolster future epidemiologic research efforts that use RDS to sample and estimate the burgeoning HIV epidemic among at-risk youth and transgender women.

Conclusions

Our findings identified important considerations for the implementation of RDS in communities of young transwomen. Qualitative data identified specific factors related to study design and implementation and community and individual characteristics that impacted participants' implementation of RDS peer referrals. Specifically, these findings identify strategies that may strengthen future RDS peer referrals and epidemiologic surveillance methods for sampling young transwomen.

Future research building on the methodology and best practices of RDS implementation is necessary to understand and improve the sampling of vulnerable and hard-to-reach minority youth in public health research. More qualitative studies examining the challenges of RDS peer referral may help to build a larger

literature base of RDS best practices. Public health practitioners and researchers can then refer and reflect on these studies to help overcome RDS peer referral challenges they may encounter. Additionally, future studies assessing webRDS in comparison with traditional in-person RDS could reveal important findings about when and for which population Internet technologies play a critical role in reaching.

Acknowledgments

The authors would like to thank all participants in the SHINE study. Award Number R01MH095598 from the National Institute of Mental Health funded this study. The lead author was supported by Award Number T32AA007240, Graduate Research Training on Alcohol Problems, from the National Institute on Alcohol Abuse and Alcoholism. Award Number R25MH097591 from the National Institute of Mental Health supported the second author. The content is solely the responsibility of the authors and does not necessarily represent the official view of the National Institute on Alcohol Abuse and Alcoholism, the National Institute of Mental Health, or the National Institutes of Health.

Conflicts of Interest

None declared.

References

1. Magnani R, Sabin K, Saidel T, Heckathorn D. Review of sampling hard-to-reach and hidden populations for HIV surveillance. AIDS 2005 May;19 Suppl 2:S67-S72. [Medline: 15930843]
2. Faugier J, Sargeant M. Sampling hard to reach populations. J Adv Nurs 1997 Oct;26(4):790-797. [Medline: 9354993]
3. 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]
4. Heckathorn D. Respondent-driven sampling: A new approach to the study of hidden populations. Social Problems 1997 May;44(2):174-199. [doi: 10.2307/3096941]
5. Baral SD, Poteat T, Strömdahl S, Wirtz AL, Guadamuz TE, Beyrer C. Worldwide burden of HIV in transgender women: A systematic review and meta-analysis. Lancet Infect Dis 2013 Mar;13(3):214-222. [doi: 10.1016/S1473-3099(12)70315-8] [Medline: 23260128]
6. Clements-Nolle K, Marx R, Guzman R, Katz M. HIV prevalence, risk behaviors, health care use, and mental health status of transgender persons: Implications for public health intervention. Am J Public Health 2001 Jun;91(6):915-921. [Medline: 11392934]
7. Silva-Santisteban A, Raymond HF, Salazar X, Villayzan J, Leon S, McFarland W, et al. Understanding the HIV/AIDS epidemic in transgender women of Lima, Peru: Results from a sero-epidemiologic study using respondent driven sampling. AIDS Behav 2012 May;16(4):872-881. [doi: 10.1007/s10461-011-0053-5] [Medline: 21983694]
8. Rotondi NK, Bauer GR, Scanlon K, Kaay M, Travers R, Travers A. Nonprescribed hormone use and self-performed surgeries: "Do-it-yourself" transitions in transgender communities in Ontario, Canada. Am J Public Health 2013 Oct;103(10):1830-1836. [doi: 10.2105/AJPH.2013.301348] [Medline: 23948009]
9. Rapues J, Wilson EC, Packer T, Colfax GN, Raymond HF. Correlates of HIV infection among transfemales, San Francisco, 2010: Results from a respondent-driven sampling study. Am J Public Health 2013 Aug;103(8):1485-1492. [doi: 10.2105/AJPH.2012.301109] [Medline: 23763398]
10. Wilson EC, Garofalo R, Harris RD, Herrick A, Martinez M, Martinez J, Transgender Advisory Committeethe Adolescent Medicine Trials Network for HIV/AIDS Interventions. Transgender female youth and sex work: HIV risk and a comparison of life factors related to engagement in sex work. AIDS Behav 2009 Oct;13(5):902-913 [FREE Full text] [doi: 10.1007/s10461-008-9508-8] [Medline: 19199022]
11. Bauermeister JA, Zimmerman MA, Johns MM, Glowacki P, Stoddard S, Volz E. Innovative recruitment using online networks: Lessons learned from an online study of alcohol and other drug use utilizing a web-based, respondent-driven sampling (webRDS) strategy. J Stud Alcohol Drugs 2012 Sep;73(5):834-838. [Medline: 22846248]
12. Zembe YZ, Townsend L, Thorson A, Ekström AM. Predictors of inconsistent condom use among a hard to reach population of young women with multiple sexual partners in peri-urban South Africa. PLoS One 2012;7(12):e51998 [FREE Full text] [doi: 10.1371/journal.pone.0051998] [Medline: 23284847]
13. Kuhns LM, Kwon S, Ryan DT, Garofalo R, Phillips G, Mustanski BS. Evaluation of respondent-driven sampling in a study of urban young men who have sex with men. J Urban Health 2015 Feb;92(1):151-167. [doi: 10.1007/s11524-014-9897-0] [Medline: 25128301]
14. Phillips G, Kuhns LM, Garofalo R, Mustanski B. Do recruitment patterns of young men who have sex with men (YMSM) recruited through respondent-driven sampling (RDS) violate assumptions? J Epidemiol Community Health 2014 Dec;68(12):1207-1212 [FREE Full text] [doi: 10.1136/jech-2014-204206] [Medline: 25086159]

15. McCreesh N, Frost Simon D W, Seeley J, Katongole J, Tarsh MN, Ndunguse R, et al. Evaluation of respondent-driven sampling. Epidemiology 2012 Jan;23(1):138-147 [FREE Full text] [doi: 10.1097/EDE.0b013e31823ac17c] [Medline: 22157309]
16. United NE. United Nations Educational, S.a.C.O. What do we mean by "youth"? URL: http://www.unesco.org/new/en/social-and-human-sciences/themes/youth/youth-definition/ [accessed 2015-08-14] [WebCite Cache ID 6ambxNvQu]
17. Corbin J, Strauss A. Basics of qualitative research: Techniques and procedures for developing grounded theory. USA: Sage Publications, Inc; 2008.
18. Truong HM, Grasso M, Chen Y, Kellogg TA, Robertson T, Curotto A, et al. Balancing theory and practice in respondent-driven sampling: A case study of innovations developed to overcome recruitment challenges. PLoS One 2013;8(8):e70344 [FREE Full text] [doi: 10.1371/journal.pone.0070344] [Medline: 23990901]
19. Hathaway AD, Hyshka E, Erickson PG, Asbridge M, Brochu S, Cousineau M, et al. Whither RDS? An investigation of respondent driven sampling as a method of recruiting mainstream marijuana users. Harm Reduct J 2010;7:15 [FREE Full text] [doi: 10.1186/1477-7517-7-15] [Medline: 20618944]
20. No authors listed. Erratum for PMID 21180585. Therap Adv Gastroenterol 2012 Sep;5(5):371 [FREE Full text] [doi: 10.1177/1756283X10363751] [Medline: 22973420]
21. Sadasivam RS, Volz EM, Kinney RL, Rao SR, Houston TK. Share2Quit: Web-based peer-driven referrals for smoking cessation. JMIR Res Protoc 2013;2(2):e37 [FREE Full text] [doi: 10.2196/resprot.2786] [Medline: 24067329]
22. Bengtsson L, Lu X, Nguyen QC, Camitz M, Hoang NL, Nguyen TA, et al. Implementation of web-based respondent-driven sampling among men who have sex with men in Vietnam. PLoS One 2012;7(11):e49417 [FREE Full text] [doi: 10.1371/journal.pone.0049417] [Medline: 23152902]
23. Stein ML, van Steenbergen Jim E, Chanyasanha C, Tipayamongkholgul M, Buskens V, van der Heijden Peter G M, Kretzschmar Mirjam E E. Online respondent-driven sampling for studying contact patterns relevant for the spread of close-contact pathogens: A pilot study in Thailand. PLoS One 2014;9(1):e85256 [FREE Full text] [doi: 10.1371/journal.pone.0085256] [Medline: 24416371]
24. Wendler DS. Assent in paediatric research: Theoretical and practical considerations. J Med Ethics 2006 Apr;32(4):229-234 [FREE Full text] [doi: 10.1136/jme.2004.011114] [Medline: 16574878]
25. Morrison-Beedy D, Carey MP, Aronowitz T, Mkandawire L, Dyne J. Adolescents' input on the development of an HIV risk reduction intervention. J Assoc Nurses AIDS Care 2002;13(1):21-27. [Medline: 11828857]
26. Ziff MA, Willard N, Harper GW, Bangi AK, Johnson J, Ellen JM. Connect to protect researcher-community partnerships: Assessing change in successful collaboration factors over time. Glob J Community Psychol Pract 2010 Jan;1(1):32-39 [FREE Full text] [Medline: 21152354]
27. DiClemente RJ, Ruiz MS, Sales JM. Barriers to adolescents' participation in HIV biomedical prevention research. J Acquir Immune Defic Syndr 2010 Jul;54 Suppl 1:S12-S17 [FREE Full text] [doi: 10.1097/QAI.0b013e3181e1e2c0] [Medline: 20571418]
28. Brawner BM, Volpe EM, Stewart JM, Gomes MM. Attitudes and beliefs toward biobehavioural research participation: Voices and concerns of urban adolescent females receiving outpatient mental health treatment. Ann Hum Biol 2013;40(6):485-495. [doi: 10.3109/03014460.2013.806590] [Medline: 23822716]
29. Valaitis R. "They don't trust us; we're just kids": Views about community from predominantly female inner city youth. Health Care Women Int 2002;23(3):248-266. [doi: 10.1080/073993302317346325] [Medline: 12003503]
30. Miller MK, Wickliffe J, Jahnke S, Linebarger JS, Dowd D. Accessing general and sexual healthcare: Experiences of urban youth. Vulnerable Child Youth Stud 2014 Jul 1;9(3):279-290 [FREE Full text] [doi: 10.1080/17450128.2014.925170] [Medline: 25101138]

Abbreviations

HIV: human immunodeficiency virus
IRB: Institutional Review Board
RDS: respondent-driven sampling
webRDS: Web adaptations of RDS
YMSM: young men who have sex with men

Edited by T Sanchez, G Eysenbach; submitted 23.04.15; peer-reviewed by T Finlayson, R Britt; comments to author 29.06.15; revised version received 13.07.15; accepted 21.07.15; published 25.08.15.

Please cite as:
Arayasirikul S, Cai X, Wilson EC
A Qualitative Examination of Respondent-Driven Sampling (RDS) Peer Referral Challenges Among Young Transwomen in the San Francisco Bay Area
JMIR Public Health Surveill 2015;1(2):e9
URL: http://publichealth.jmir.org/2015/2/e9/
doi:10.2196/publichealth.4573
PMID:27227143

Original Paper

Texting-Based Reporting of Adverse Drug Reactions to Ensure Patient Safety: A Feasibility Study

Godofreda Vergeire-Dalmacion[1,2,3], MSc, MD, MHPEd; Nina T Castillo-Carandang[2,3], MSc; Noel R Juban[2,3], MSc, MD; Maria Lourdes Amarillo[2,3], MPH; Maria Pamela Tagle[2,3], BSc; Emmanuel S Baja[2,3], MSc, ScD

[1]College of Medicine, Department of Pharmacology and Toxicology, University of the Philippines-Manila, Manila, Philippines

[2]College of Medicine, Department of Clinical Epidemiology, University of the Philippines-Manila, Manila, Philippines

[3]National Institutes of Health, Institute of Clinical Epidemiology, University of the Philippines-Manila, Manila, Philippines

Corresponding Author:
Godofreda Vergeire-Dalmacion, MSc, MD, MHPEd
College of Medicine
Department of Pharmacology and Toxicology
University of the Philippines-Manila
Department of Clinical Epidemiology Room 103
Paz Mendoza Building, Pedro Gil Street, Ermita
Manila, 1000
Philippines
Phone: 63 2 525 4098
Fax: 63 2 525 4098
Email: jody.dalmacion@gmail.com

Abstract

Background: Paper-based adverse drug reaction (ADR) reporting has been in practice for more than 6 decades. Health professionals remain the primary source of reports, while the value of patients' reporting is yet unclear. With the increasing popularity of using electronic gadgets in health, it is expected that the electronic transmission of reports will become the norm within a few years.

Objective: The aims of this study are to investigate whether short messaging service or texting can provide an alternative or supplemental method for ADR reporting given the increasing role of mobile phones in health care monitoring; to determine the usefulness of texting in addition to paper-based reporting of ADRs by resident physicians; and to describe the barriers to ADR reporting and estimate the cost for setting up and maintaining a texting-computer reporting system.

Methods: This was a pre-post cross-sectional study that measured the number of ADRs texted by 51 resident physicians for 12 months from the Department of Obstetrics and Gynecology and the Department of Adult Medicine of a tertiary government hospital in Manila, Philippines, with 1350-bed capacity. Reports were captured by a texting-computer reporting system. Prior to its implementation, key informant interview and focus group discussion were conducted. Baseline information and practice on the existing paper-based reporting system were culled from the records of the hospital's Pharmacy and Therapeutics Committee. A postintervention survey questionnaire was administered at the end of 12 months.

Results: Only 3 ADRs were texted by 51 resident physicians in 12 months (reporting rate 3/51 or 6%). By contrast, 240 ADRs from the paper-based reporting system from 848 resident physicians of the study hospital were collected and tabulated (reporting rate 240/848 or 28.3%). Texting ADRs was not efficient because of power interruption, competition with the existing paper-based reporting system, and unforeseen expiration of prepaid text loads/credits. The 3 ADRs texted were a report of vivid dreams and nightmares, a report of disturbing dreams and memory lapses, both of which were due to montelukast use, and a report of hepatitis from an isoniazid/rifampicin fixed-dose combination. Nineteen of 51 resident physicians (37%) registered in the reporting system responded to the postintervention survey. The most common reasons for not reporting ADRs were no adverse reaction identified 11/19 (58%) and restrictive reporting syntax 4/19 (21%). All doctors preferred a free form of reporting. The direct cost of the texting-based reporting system was calculated to be US $5581.40 and the indirect cost was US $9989.40. The total cost for texting-based ADR reporting system for 12 months was US $15,570.79.

Conclusions: Reporting of ADRs via texting could be lower compared with an existing ADR paper-based system. Problems of Internet connectivity, reporting syntax, and expiration and reliability of text loads/credits should be addressed while implementing a text-based ADR reporting system in a developing country.

(JMIR Public Health Surveill 2015;1(2):e12) doi:10.2196/publichealth.4605

KEYWORDS

adverse drug reactions; pharmacovigilance; postmarketing; spontaneous reporting; texting

Introduction

Toward the early 1970s, pharmacovigilance evolved as a critical field in drug development and regulation after the drug thalidomide showed serious adverse reactions in humans but not in animals. As an offshoot of this disastrous incident, in 1968, the World Health Organization (WHO) established the Program for International Drug Monitoring in Uppsala, Sweden [1]. The program created an adverse drug reaction (ADR) monitoring system based on spontaneous reporting by health care professionals. In the early 1980s, regulators, pharmaceutical manufacturers, and physicians realized that prolonging the approval of new drugs is as harmful as allowing marketing of drugs without postmarketing surveillance. Regulators also realized that rare ADRs, effects from drug-drug interactions, drug-disease interaction, and self-medication toxicities can only be elucidated in the real world of drug treatment rather than in clinical trials. This paved the way to the science of pharmacoepidemiology and the practice of pharmacovigilance.

To promote and encourage postmarketing surveillance, various strategies, regulatory policies, and even laws were created to facilitate the reporting of suspected ADRs. The major task for assuring drug safety was given to the pharmaceutical companies. However, this resulted in a conflict of interest and a restrained willingness to pass judgment on a drug's culpability. A case in point is cerivastatin. After the release of cerivastatin on February 18, 1998, the dataset of ADRs grew, but important analyses of these data remained internal to Bayer Corporation [2]. Bayer modified the label of cerivastatin 5 times during the 3 years the drug was available to try to improve its safety before ultimately withdrawing it from the market. Reports of ADRs are often inadequately recorded or defined. This was proven by Loke and Derry in 2001 in their systematic review of the reporting and recording of ADRs in 185 randomized clinical trials. They found that 25 of the 185 trials (13.5%) did not mention anything about ADRs [3]. When ADRs such as clinical events or patient symptoms were mentioned in the reports, details on how they had been recorded were given in only 14 of 95 (15%) and 18 of 104 (17%) trials, respectively [3].

After 4 decades, postmarketing surveillance of newly marketed drugs has become a vital step in drug development, demanding the same attention and rigors as the other steps. It has also led to the science of pharmacovigilance, which is defined as "activities relating to the detection, assessment, understanding and prevention of adverse effects or any other drug-related problem." Together with the WHO Collaborating Centre for International Drug Monitoring, Uppsala, WHO promotes pharmacovigilance at the country level. At the end of 2010, 134 countries have become part of the WHO pharmacovigilance program. Recently, pharmacovigilance has become synonymous with patient safety and care in relation to the use of medicines [1]. Although spontaneous ADR reporting is the mainstay of safety evaluation in the postapproval phase, a high level of unquantifiable underreporting by doctors remains an insurmountable problem of the system [4].

Another systematic review involving 37 studies across 12 countries was carried out in 2006 to numerically estimate the underreporting of ADRs using spontaneous reporting. The median underreporting rate was 94% (interquartile range 82-98%) [5]. There was no significant difference in the median underreporting rates calculated for general practice and hospital-based studies [5]. Despite more than 5 decades of spontaneous reporting, underreporting still remains a major drawback. Health professionals remain the primary source of reports, while the value of patients' reporting is yet unclear. With the increasing popularity of using electronic gadgets in health, it is expected that the electronic transmission of reports will become the norm within a few years [6]. Ines in 1986 [7] published the earliest body of evidence on ADR reporting in the Philippines. This aforementioned study revealed that ADR reporting established in 1967 by the Philippine Medical Association has not been successful in gathering information on ADRs for the last 20 years of its existence. The author proposed in her study a change in the attitudes of patients and doctors toward ADRs and the adoption of newer strategy of reporting. Furthermore, a confidential and reliable mechanism of review and assessing the reports was recommended.

In the era of computers, cyberspace, and rapid connectivity, it is reasonable to explore the potential use of short messaging service, more commonly known as "texting," in reporting ADRs. The use of mobile phones in health delivery services and care is not new. In 2008, mobile phone texting for pharmaceutical care in a hospital was implemented in China [8]. The system was called the "Mobile Pharmacy Service System." The text messages sent by the system to patients consisted of the following: (1) reminders about medication from the day following discharge, (2) practical information about medicines, and (3) information about ADRs. Baron et al [9] published in 2013 a pilot study on the use of mobile phone-based tools for adverse event notifications after a vaccination program in Cambodia. A total of 184 patients from the study were texted for their clinical status 48 hours after their vaccinations. Fifty-two (28.3%) did not reply but 101 (54.9%) sent an immediate text response, and 31 (16.8%) sent a text reply after additional prompting. The study concluded that texting can also be a useful tool for notification by patients or health users in Cambodia, especially in an urban setting [9]. Local information on the texting activity of the Philippines in 2007 revealed that

roughly 50 million are registered text message users and a staggering average of 195 texts were sent per user per month in the same year [10].

Our study aimed to test the usefulness of texting in conjunction with paper-based reporting of ADRs by resident physicians and to describe the barriers to ADR reporting; to determine the rate of texts reporting ADRs by resident physicians of 2 departments in a government tertiary care hospital; to determine the most commonly reported ADRs and the suspect drugs; and to estimate direct and indirect costs of texting for reporting ADRs.

Methods

Study Site and Design

This was a pre-post cross-sectional study with text-based ADR reporting system as the intervention. The study covered a period of 12 months from April 2011 to March 2012 excluding 3 months of preparatory work on the computer-texting system. The University of the Philippines-Philippine General Hospital (UP-PGH), a major government tertiary care hospital with 1350-bed capacity in the city of Manila, was chosen as the study site. The selection of the UP-PGH was based on the following criteria: presence of an active Pharmacy and Therapeutics Committee, high prescriptions based on annual patient admissions, a roster of health providers that remains relatively constant every year, and a higher probability of tracking down any patient reported to have an ADR. Two purposively selected study sites in the hospital, namely the Department of Adult Medicine and the Department of Obstetrics and Gynecology, were included in the study. The choice of these 2 departments was based on the expected high rate of prescribing of medications. Information materials and registration forms of the texting-based ADR reporting were disseminated 2 weeks prior to the launch of the project. Fifty-one resident physicians from the Department of Adult Medicine and the Department of Obstetrics and Gynecology signed the informed consent and were subsequently registered to the texting-based reporting system. Furthermore, information about the resident physician was entered in their registration directory, including their mobile phone numbers, age, gender, department, and year of residency. Key informant interview, review of hospital records, and focus group discussion (FGD) were conducted prior to the implementation of the texting-based reporting. After 12 months, a 2-paged survey questionnaire was administered to the 51 resident physician registered in the texting-based reporting to obtain their perception toward the new reporting strategy. In the same session, respondents were further probed about some of their replies to the questionnaire, which needed clarification. Moreover, during the entire study period, ADRs from the paper-based reporting system from all of the 848 resident physicians of the study hospital were collected and tabulated. In addition, direct and indirect costs were estimated using local currency and subsequently converted into US dollars.

Creating the Internet-Based Reporting System With a Mobile Phone Interface

An information technology (IT) consultant was commissioned to create the texting-computer reporting system. It took 2 IT providers and almost 12 months of configuring the computer with the texting interface before the final form was ready for installation. The system utilized all the 3 local cellular phone companies that provide texting services in the country. All text messages received by any of the 3 mobile phone companies were automatically captured and sent to the database stored in the system. Anyone sending a report to the system was automatically acknowledged by a text message. The reporting system was given the acronym “DIMES,” or the “Drug Information and Monitoring Event for Safety.” A reporting syntax was required to be followed by the reporters for their texts to be accepted by the system. The database in the computer was also structured following the same order. The system was accessible 24 hours a day and 7 days a week and was configured to set an alert if there was a cluster of similar ADRs from one or more drugs or one drug repeatedly reported for several ADRs entering the system. The system administrator regularly reviewed the database for any signals or technical glitches. The research team used the Naranjo algorithm to produce a causality determination guideline for the study group [11]. The team also drafted a user’s manual to help the study group understand the technical features of the texting system and troubleshoot glitches. For the first 5 months of implementation, advisories on recent drug withdrawals and emerging profiles of new ADRs were sent to all texting-based registrants. These advisories were intended to serve as prompts indicating that the system was active and functioning.

Results

Result of the Key Informant Interview

Our key informant was a former director of the Philippine Food and Drug Administration (FDA). According to her, the Philippines was one of the earliest countries in Southeast Asia to become a member of the WHO Collaborating Centre for International Drug Monitoring in February 1995. The administration regularly sends ADR reports generated from its paper-based reporting system to Uppsala, Sweden. Initially, the members of the National Adverse Drug Reaction Advisory Committee assessed the reports for possible causality using global introspection. Our key informant directly supervised this committee’s operation for many years before her retirement. According to the key informant, the committee disbanded in 2006, after which almost no ADR reports were recorded. In 2007, the Philippine FDA launched a WHO-supported pharmacovigilance strategy, “Bantay Gamot,” or “Drug Watch.” Drug Watch is a paper-based consumer-reporting scheme, which continues to effectively receive reports on the substandard quality of drugs. In 2010, the Philippine FDA collaborated with Department of Health Information Management System to establish an online ADR reporting system. However, many health professionals and pharmacovigilance officers from drug companies reported difficulty with opening the site, a major drawback to its effectiveness. In 2011, the Philippine FDA received 2032 ADR reports, of which 691 were sent to Uppsala, Sweden. The UP-PGH’s Pharmacy and Drug Committee was one of the regular contributors to the Philippines FDA.

Result of the Focus Group Discussion

Ten resident physicians from the 2 study sites participated in the FGD. A summary of the findings from the discussion is shown in Textbox 1.

Textbox 1. Results of the focus group discussion describing the knowledge, attitude, and practice of resident physicians toward drug prescribing and adverse drug reactions.

Knowledge of drug prescribing

- The most commonly prescribed drugs were antihistamines, steroids, chemotherapeutics, ferrous sulfates, contraceptives, metronidazole, insulin, and analogues.
- The usual number of drugs taken per patient is between 2 and 3.
- 75% of prescriptions are orally administered.

Knowledge of ADRs

- Overall, ADRs are very rare and are usually managed on an outpatient basis.
- Common types of reactions are anticipated side effects (9 of 10), and unforeseen events (1 of 10).
- ADRs from contrast media should be included as source of ADRs.
- The issue of therapeutic failure due to substandard generics should be included.

Physicians' knowledge and attitude toward patients reporting ADRs

- Patients already report adverse reactions to their doctors using text messaging mostly about their follow-ups or consulting earlier than scheduled.
- Patients usually consult for drug-related allergies.
- Female patients (including female relatives of male patients) report ADRs more than males.
- Doctors believe that patients are not in a good position to report ADRs.
- Physicians tend to only ask about known side effects.

Attitude toward reporting of ADRs

- Possible barriers to reporting ADRs include loss of the plan/texting ability due to payment/credit expiration, no mobile phone, no transportation money, paperwork, no signal, rejection of texts because they are too long.
- Factors facilitating adverse drug response reporting are good rapport with the doctor, clear symptoms, giving the responsibility to female relatives, use of a specific telecommunications provider, and providing feedback to the doctor.
- Participation concerns regarding the texting-based system include availability of journal articles as support to alerts, assurance that the texting-based system will not replace the doctor or pit one doctor against another.

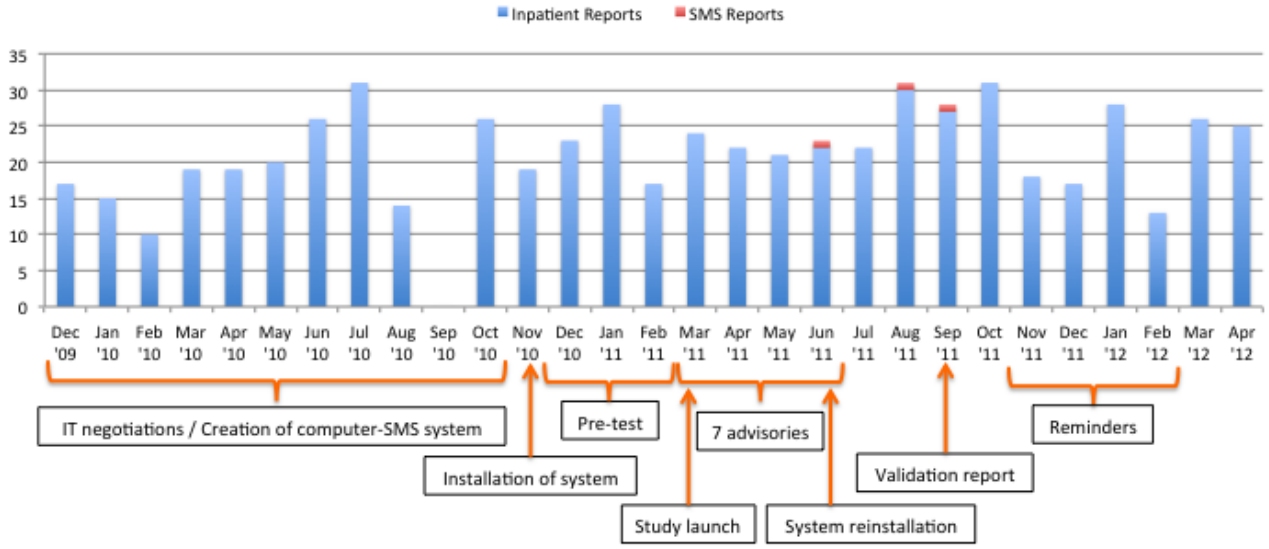
All resident physicians were receptive to using texting to report ADRs. They also suggested using texts to send clinical evidence from journals. Not surprisingly, all the participants had used texting to remind patients of their clinic visits and to reply to patients' queries about dosing and drug administration. Most agreed that reporting is time consuming and all reporters expect to receive feedback on their reports.

Results of Texting-Based and Paper-Based Reporting

Figure 1 shows the ADR results of both the texting-based and paper-based reporting systems in the hospital. A total of 277 ADR reports from the paper-based system of the UP-PGH were recorded the year before the implementation of the texting-based reporting system. No paper-based ADR report was available in September 2010. The reports came from the 848 resident physicians of all departments and sections of the hospital. The total number of resident physicians in the entire hospital is basically the same through the years due to the fixed number of positions in each department and section of the hospital. The reporting rate from the paper-based system prior to the implementation of the texting-based system was 32.7% (277/848).

During the implementation of the texting-based system, a total of 240 ADR reports were recorded from the paper-based system giving a reporting rate of 28.3% (240/848). However, the texting-based reporting system obtained only 3 ADR reports from the 51 resident physicians of the 2 preselected specialty departments of the study hospital, which translated to a reporting rate of only 6% (3/51). Most of the ADRs received by the hospital from the paper-based reports were liver dysfunctions and jaundice from antituberculosis drugs and reports of allergies. Conversely, ADR reports from texting-based system consisted of one report of vivid dreams and nightmares and one other report of disturbing dreams and memory lapses both from montelukast. There was also one report of hepatitis from an isoniazid/rifampicin fixed-dose combination.

Figure 1. Timeline of the study before and immediately after the implementation of the texting-based reporting system and the adverse drug response reports received during this period. SMS: short messaging service.



Limitations of the Texting-Computer Hardware of the Reporting System

Numerous technical glitches were encountered, mostly brought about by fluctuating mobile phone connectivity, frequent power interruptions, and insufficient phone loads/credits. Unfortunately, these episodes of disconnection were not observed during the 2-3 months of pretesting the system; they only became evident toward the middle of the study. There were no explanations for the absence of cellular connectivity on certain days. In addition, prepaid loads/credits of the phones were already used up even 2-3 weeks before their expiration dates. To decrease the likelihood of missed reports, 1 cellular company, which was not the service provider of any of the 51 registered resident physicians, was removed from the system. The removal was done to decrease the number of phones that needed to be reloaded. Expiration of text loads/credits could have been avoided had the system been enrolled in a postpaid plan. A postpaid plan entitles the subscriber to a 4-digit mobile phone number that is easily remembered. Unfortunately, a subscriber needs to have 20,000 Philippine pesos (PhP) worth of text messages or calls per month to qualify for a postpaid plan.

Postintervention Activity

Table 1 summarized the results from the postinterventions survey among resident physicians who registered in the texting-based system. Only 37% (19/51) submitted the completed questionnaire. The major reason for not reporting ADR was the absence of an identifiable ADR (11/19; 58%) followed by the restrictive reporting syntax (4/19; 21%).

Table 1. Postintervention survey among resident physicians registered in the text reporting on their perception toward the adverse drug reaction texting-based reporting (N=19).

Survey questions		n (%)
Resident physicians who received advisories		
	Yes	12 (63)
	No	5 (26)
	No reply	2 (11)
Reasons for not reporting via text		
	No identified ADRs	14 (74)
	Reported through paper based	1 (5)
	No answer	4 (21)
Reasons for not reporting ADR via any reporting method		
	No identifiable ADRs	11 (58)
	Constraining text syntax	4 (21)
	Presence of paper-based	4 (21)
	Lack of time	1 (5)
	Lack of perceived need	1 (5)
	Fear of litigation	0 (0)

Cost Estimates

The direct and indirect costs of the text-based reporting system were computed. For the direct cost, the cost of the computer, 3 mobile phones, and a one-time cost of the subscriber identification module (SIM) cards were included in the computation. The total direct cost was 240,000 PhP or US $5581.40 at an exchange rate of US $1 to 43.00 PhP. The indirect cost included the prepaid phone loads for 12 month at 19,544.00 PhP (US $454.51), the IT consultancy fee of 50,000.00 PhP (US $1162.79), and the honorarium of the system administrator of 360,000.00 PhP (US $8372.09) for 12 months. The total indirect cost was 429,544.00 PhP (US $9989.40). Adding the direct and indirect costs, the total estimated cost was 669,544.00 PhP or US $15,570.79 to establish and run the texting-based reporting for 1 year.

Discussion

Preliminary Findings

The Philippine FDA handles the Adverse Drug Reaction Spontaneous Reporting System in the Philippines. According to our key informant, underreporting of ADRs has perennially hampered the reporting system in the Philippines. Historically, the system has received only 1000-3000 reports per year. Yet, in 2003, drug sales from the largest drug retailer in the Philippines amounted to nearly 43 billion PhP [12]. ADR reporting to vary from 7% of all hospital admissions in the UK to 13% of all admissions in medical clinics in Sweden [13]. In New Zealand, 12.9% of all hospital admissions were due to adverse drug events [13,14]. To improve the detection of previously unknown serious ADRs, the US FDA introduced the MEDWATCH program in 1993 [13]. Approximately 1 year from its introduction, the number and quality of ADR reports to the FDA increased. However, this rise was attributed to increased reporting from pharmacists. Physician reports declined slightly during this period. Although the medical literature is rich on studies about ADRs, there is none on the use of texting as a strategy for reporting them. To our knowledge, our study is the first look at texting as an alternative method of reporting ADR. In addition, our study was conducted on the premise that reporting of ADRs can be improved with the use of a ubiquitous and popular communication technology such as texting. However, our study found no increase in the number of ADR reports using texting. Our study also revealed that texting was highly dependent on a reliable telecommunication services, which entails a relatively large amount of monthly postpaid plan of 20,000.00 PhP or US $435.00. In addition, efficient syntax reporting plays a critical role in reporting ADR in texting-based reporting system.

The reporting rate from the texting-based system (3/51; 6%) of our study was absolutely lower compared with the reporting rate from the paper-based system (240/848; 28.3%). Our finding should be regarded with caution because the calculation of the reporting rate for both the texting-based and paper-based systems used a different total number of resident physicians (51 vs 848). The texting-based system only received ADR reports from 2 departments of the study hospital, whereas the paper-based system covered all specialty departments and sections of the hospital. These 2 departments were purposively selected because of their higher use of medications compared with purely surgical departments. The reporting rates were used simply to describe the state of both reporting system. Nevertheless, more studies are needed to verify and validate the efficiency of a texting-based reporting system as an alternative system in ADR reporting.

Brewer and Colditz in 1999 [15] observed that spontaneous reporting systems could be effective in revealing unusual and rare events that occur with the use of medications. However, they showed that spontaneous reporting systems were not reliable for detecting ADRs occurring far from the time of intake or in a population not commonly exposed to the drug. Brewer and Colditz [15] recommended the use of other methods, such as clinical trial data, medical records, and computerized databases of medication users and nonusers to complement spontaneous reporting. Huang et al [16] reviewed in 2014 the postmarketing drug surveillance for adverse drug events worldwide. It showed 2 systems of surveillance in the UK using administrative claims or electronic medical records and most pharmacovigilance being conducted on behalf of a regulatory agency. To access existing data, either a common data model or a centralized model could be used. Aside from studying existing databases as data sources for detecting ADRs, methods for reporting ADR such as texting were explored by our study. However, results from our study showed that texting could be unreliable due to consumable phone loads/credits and an unpredictable power supply. Nonetheless, these obstacles may be less common in more economically developed countries.

Resident physicians were probed about their reply of "no identified ADR" as a reason for not reporting. About a third of the respondents did not find it necessary to report known and accepted adverse reactions to a suspect drug. The rest really admitted to not identifying any ADR during the 12-month study period. The observation that few ADR reports were due to "unrecognized adverse events" was likewise reported in a study by Hartigan-Go in 2002 [17]. Hartigan-Go [17] reported that sometimes the adverse event is misconstrued as part of the healing action. This suggests that physicians and other health professional prescribing medications should be regularly educated on the principles and rationale of pharmacovigilance and be reminded on how important it is to quantify even known ADRs to drugs in their local clinical setting. Contrary to our expectations, fear of litigation was never considered a deterrent to reporting.

However, there is still room for studying the use of texting for reporting ADR by other health professionals, such as nurses and clinical pharmacists, provided the observed obstacles and problems are resolved. More people reporting will help generate the 20,000.00 PhP required for the postpaid plan that will ensure uninterrupted Internet service. Lastly, texting might be best suited for timely dissemination of drug information bulletins, drug advisories, or as a tool for reminding patients on their follow-up visits. A study by Kew in 2010 [18] found that texting via mobile phone was an effective method for collecting weekly symptom reports during a clinical trial, reminding trial patients to attend face-to-face visits and completing more complex paper-based evaluation.

Conclusions

In summary, the reporting rate of ADRs using texting-based ADR reporting system may be lower compared with the paper-based ADR reporting system. Unreliable telecommunication services, frequent electrical interruptions, reporting syntax, and expiring prepaid loads/credits should be addressed when setting up a texting-based ADR reporting system in a developing country.

Conflicts of Interest

None declared.

References

1. World Health Organization-Uppsala Monitoring Center. Uppsala Monitoring Center: Home. Uppsala, Sweden: Uppsala Monitoring Center URL: http://www.who-umc.org/ [accessed 2015-11-15] [WebCite Cache ID 6bVx7sfge]
2. Psaty BM, Furberg CD, Ray WA, Weiss NS. Potential for conflict of interest in the evaluation of suspected adverse drug reactions: Use of cerivastatin and risk of rhabdomyolysis. JAMA 2004 Dec 1;292(21):2622-2631. [doi: 10.1001/jama.292.21.2622] [Medline: 15572720]
3. Loke YK, Derry S. Reporting of adverse drug reactions in randomised controlled trials: A systematic survey. BMC Clin Pharmacol 2001;1:3 [FREE Full text] [Medline: 11591227]
4. Fanelli R, Di GM, Di MM, Loperfido F, De MD. Interference of unstable intraventricular conduction disorders in the recognition of anterior necrosis. Boll Soc Ital Cardiol 1979;24(10):1149-1155. [Medline: 261900]
5. Hazell L, Shakir SA. Under-reporting of adverse drug reactions : A systematic review. Drug Saf 2006;29(5):385-396. [Medline: 16689555]
6. Coulter DM. The New Zealand Intensive Medicines Monitoring Programme. Pharmacoepidemiol Drug Saf 1998 Mar;7(2):79-90. [doi: 10.1002/(SICI)1099-1557(199803/04)7:2<79::AID-PDS330>3.0.CO;2-1] [Medline: 15073731]
7. Ines CE. Adverse drug reaction reporting. Philippine experience. Med Toxicol 1986;1 Suppl 1:99-104. [Medline: 3821435]
8. Mao Y, Zhang Y, Zhai S. Mobile phone text messaging for pharmaceutical care in a hospital in China. J Telemed Telecare 2008;14(8):410-414. [doi: 10.1258/jtt.2008.080406] [Medline: 19047450]
9. Baron S, Goutard F, Nguon K, Tarantola A. Use of a text message-based pharmacovigilance tool in Cambodia: Pilot study. J Med Internet Res 2013;15(4):e68 [FREE Full text] [doi: 10.2196/jmir.2477] [Medline: 23591700]
10. J@ps [Hubpages]. The Philippines Is Still the Undisputed Text Capital of the World. 2011 Aug 9. URL: http://hubpages.com/travel/The-Philippines-is-Still-the-Undisputed-Text-Capital-of-the-World [accessed 2015-11-16] [WebCite Cache ID 6d4q1qVnr]

11. Naranjo CA, Busto U, Sellers EM, Sandor P, Ruiz I, Roberts EA, et al. A method for estimating the probability of adverse drug reactions. Clin Pharmacol Ther 1981 Aug;30(2):239-245. [Medline: 7249508]
12. Reyes CM, Lavado R, Tabuga AD, Asis RD, Datu MBG. A Profile of the Philippine Pharmaceutical Sector. 2011 May. URL: http://dirp4.pids.gov.ph/webportal/CDN/PUBLICATIONS/pidsdps1111_rev.pdf [accessed 2015-11-16] [WebCite Cache ID 6d4qsGSSO]
13. Gony M, Badie K, Sommet A, Jacquot J, Baudrin D, Gauthier P, et al. Improving adverse drug reaction reporting in hospitals: Results of the French Pharmacovigilance in Midi-Pyrénées region (PharmacoMIP) network 2-year pilot study. Drug Saf 2010 May 1;33(5):409-416. [doi: 10.2165/11319170-000000000-00000] [Medline: 20397740]
14. Piazza-Hepp TD, Kennedy DL. Reporting of adverse events to MedWatch. Am J Health Syst Pharm 1995 Jul 1;52(13):1436-1439. [Medline: 7671043]
15. Brewer T, Colditz GA. Postmarketing surveillance and adverse drug reactions: Current perspectives and future needs. JAMA 1999 Mar 3;281(9):824-829. [Medline: 10071004]
16. Huang YL, Moon J, Segal JB. A comparison of active adverse event surveillance systems worldwide. Drug Saf 2014 Aug;37(8):581-596 [FREE Full text] [doi: 10.1007/s40264-014-0194-3] [Medline: 25022829]
17. Hartigan-Go K. Developing a pharmacovigilance system in the Philippines, a country of diverse culture and strong traditional medicine background. Toxicology 2002 Dec 27;181-182:103-107. [Medline: 12505293]
18. Kew S. Text messaging: An innovative method of data collection in medical research. BMC Res Notes 2010;3:342 [FREE Full text] [doi: 10.1186/1756-0500-3-342] [Medline: 21172018]

Abbreviations

ADR: adverse drug reaction
FDA: Food and Drug Administration
FGD: focus group discussion
IT: information technology
SIM: subscriber identification module
UP-PGH: University of the Philippines-Philippine General Hospital
WHO: World Health Organization

Edited by T Sanchez; submitted 03.05.15; peer-reviewed by C Jimeno; comments to author 07.07.15; revised version received 06.08.15; accepted 19.08.15; published 19.11.15.

Please cite as:
Vergeire-Dalmacion G, Castillo-Carandang NT, Juban NR, Amarillo ML, Tagle MP, Baja ES
Texting-Based Reporting of Adverse Drug Reactions to Ensure Patient Safety: A Feasibility Study
JMIR Public Health Surveill 2015;1(2):e12
URL: http://publichealth.jmir.org/2015/2/e12/
doi:10.2196/publichealth.4605
PMID:27227130

Original Paper

Direct-to-Patient Research: Piloting a New Approach to Understanding Drug Safety During Pregnancy

Nancy A Dreyer[1], MPH, PhD; Stella CF Blackburn[2], MBBS, MA, MSc; Shahrul Mt-Isa[3], BSc, PhD; Jonathan L Richardson[4,5], BBSc (Hons); Simon Thomas[4,5], BSc, MD, FRCP; Maja Laursen[6], MSc, PhD; Priscilla Zetstra-van der Woude[7], MSc; Anna Jamry-Dziurla[8], MSc; Valerie Hliva[9], PhD; Alison Bourke[10], BSc, MSc; Lolkje de Jong-van den Berg[7], PharmD, PhD

[1]Quintiles Real-World & Late Phase Research, Cambridge, MA, United States
[2]Quintiles Real-World & Late Phase Research, Reading, United Kingdom
[3]Imperial College, London, United Kingdom
[4]Institute of Cellular Medicine, Newcastle University, Newcastle upon Tyne, United Kingdom
[5]UK Teratology Information Service, Newcastle Hospitals, NHS Foundation Trust, Newcastle upon Tyne, United Kingdom
[6]Statens Serum Institut, Copenhagen, Denmark
[7]University of Groningen, Groningen, Netherlands
[8]Poznan University of Medical Sciences, Poznan, Poland
[9]Quintiles Real-World & Late-Phase Research, St. Prex, Switzerland
[10]IMS, London, United Kingdom

Corresponding Author:
Nancy A Dreyer, MPH, PhD
Quintiles Real-World & Late Phase Research
201 Broadway, 5th floor
Cambridge, MA, 02139
United States
Phone: 1 6177156810
Fax: 1 6172069464
Email: nancy.dreyer@quintiles.com

Abstract

Background: Little is known about the effects of human fetal exposure when a new drug is authorized unless it was specifically developed for use in pregnancy. Since many factors may contribute to adverse fetal effects, having comprehensive information about in utero exposures will enhance our ability to make correct determinations about causality.

Objective: The objective of the study was to assess the extent to which women, recruited without the intervention of health care professionals (HCPs), will provide information, suitable for research purposes, via the Internet or by phone on some potential risk factors in pregnancy.

Methods: To pilot direct-to-patient research for pharmacovigilance, we conducted a prospective, noninterventional study of medication use and lifestyle factors as part of the Pharmacoepidemiological Research on Outcomes of Therapeutics by a European ConsorTium (PROTECT) Consortium. Consenting women who self-identified as pregnant and residing in the United Kingdom (UK), Denmark (DK), The Netherlands, or Poland were recruited and could then choose to provide data every 2 or 4 weeks via the Internet or a telephonic interactive voice response system (IVRS). Self-reported drug use was compared with pharmacy register data in DK and with electronic health records in the UK.

Results: Recruited women were on average older and more highly educated than the general population. Most respondents chose a frequency of every 4 weeks (56.99%, 1177/2065). Only 29.83% (464/1555) of women with due dates occurring during the study provided information on pregnancy outcome. For those responding by Internet, over 90.00% (1915/2065) reported using >1 pregnancy-related medication, 83.34% (1721/2065) reported using >1 other medicine, and 23.53% (486/2065) reported only over-the-counter medications, not counting herbals and dietary supplements. Some respondents (7.16%, 148/2065) reported that they chose not to take a prescribed medication (mostly medicines for pain or inflammation, and for depression) and 1.30% (27/2065) reported using medicines that had been prescribed to a friend or family member (oxycodone, paracetamol, and medications for acid-related problems). Relatively few respondents reported using fish oil (4.60%, 95/2065), other dietary

supplements (1.88%, 39/2065), herbal products (7.07%, 146/2065), or homeopathic products (1.16%, 24/2065). Most medications for chronic conditions that were listed in the Danish prescription registry were also self-reported (83.3%, 145/174 agreement), with larger discrepancies for medications indicated for short-term use (54.0%, 153/283 agreement) and pregnancy-related medications (66.1%, 78/118).

Conclusions: Self-reported information on medication use as well as other potential teratogenic factors can be collected via the Internet, although recruitment costs are not insubstantial and maintaining follow-up is challenging. Direct data collection from consumers adds detail, but clinical input may be needed to fully understand patients' medical histories and capture birth outcomes.

(JMIR Public Health Surveill 2015;1(2):e22) doi:10.2196/publichealth.4939

KEYWORDS

pharmacovigilance; direct-to-patient; drug safety; validation

Introduction

Prenatal Exposure to Harmful Medications

The use of medication during pregnancy may be essential for the health of the mother, but some have the potential to cause harm to the fetus, which can be delayed in presentation. Prenatal exposure to harmful medications can be related to a variety of adverse outcomes, including congenital malformations, preterm birth, intrauterine growth restriction, spontaneous abortion, late fetal death, neonatal death, or developmental disabilities (behavioral, neurological, motor, intellectual, or sensory) that only become apparent in later infancy or childhood, or even later like the rare vaginal adenocarcinomas that were detected among young women who had prenatal exposure to diethylstilbestrol [1].

Unless the medicine is intended to treat pregnancy-specific conditions, at the time of initial authorization, information with respect to reproductive toxicity is usually only available from animal studies. Pregnant women are often excluded or discontinued from premarketing clinical trials in humans, so the safety of many drugs in pregnant women has not been established at the time of drug licensing [2]. Consequently, some important drugs are contraindicated or have special warnings because their safety during pregnancy has not been studied sufficiently. In addition, women who are concerned about conventional drug use in pregnancy may turn to alternative therapies, such as homeopathic or herbal medicines, which they perceive as being "natural" and somehow safer. Many of these alternative drugs are not regulated and information about safety in pregnancy is often lacking. Many pregnancies are unplanned and inadvertent exposure of the fetus to prescription or nonprescription medications may have already occurred by the time the woman realizes she is pregnant. Because organogenesis occurs early in pregnancy, first-trimester exposures are of particular interest as detrimental effects on the fetus may have already occurred before a pregnancy is confirmed [3,4]. Therefore, good postmarketing information on drug use and other possible risk factors and pregnancy outcome is needed, not only to provide information on which drugs may be unsafe, but very importantly on which drugs are probably safe.

The majority of data available on medications used during pregnancy and lifestyle factors are usually collected either from health care professionals, through direct patient questioning by an interviewer (frequently a midwife), or making use of prescription or dispensing records. In most situations, women tend to present for medical attention once it appears the pregnancy is viable, making it difficult to get accurate information about all medication exposures that occur early in pregnancy [5]. Using researchers to collect information is time consuming and expensive, and can only be performed at relatively infrequent times during the pregnancy, which may lead to lost information [6]. In addition, women may be reluctant to report accurate information about lifestyle behaviors already identified as being potentially harmful to a fetus, or which are in themselves illegal, in a face-to-face interaction. There is some evidence that using the Internet may overcome these issues related to collection of potentially sensitive information, for example, a study on sexually transmitted diseases using an anonymous Internet questionnaire successfully collected data on the number of sexual partners and cocaine use [7].

Purpose of Pilot Study

This pilot study was designed to explore (1) whether women would be willing to volunteer for Internet-based research without any encouragement or direct involvement of their health care providers, (2) if they would provide information prospectively on exposure to medications and other factors that may affect birth outcome, and (3) if so, to assess whether this information was complete and accurate enough to be useful for pharmacovigilance.

Methods

Study Eligibility Requirements

To be eligible for study entry, women had to be pregnant; resident in 1 of 4 countries in the European Union (Denmark, DK; The Netherlands, NL; Poland, PL; or the United Kingdom, UK); to be proficient in the predominant language of their country of residence; to have access to the Internet or a telephone; and to be of an age to provide legal consent (16 years of age in the UK, 18 years elsewhere). Ethical review and review of data protection plans were reviewed as necessary in all countries, and the data protection plan was also reviewed by the European Medicines Agency and the European Data Protection Supervisor [8]. The study was promoted using a variety of methods including Internet announcements, email to members of pregnancy clubs, flyers placed in pharmacies, and radio and television interviews.

Study Participants

Women were invited to respond by Internet or interactive voice response system (IVRS), using the predominant natural language in each of the 4 study countries. The IVRS system was developed in addition to Internet-based systems to facilitate participation regardless of Internet access, thus making the study available to women without access to computers, or the requisite computer skills. In NL and the UK, women could use either method for enrollment and informed consent, but in DK, informed consent by Internet was required even for subsequent participation by IVRS. In PL, the IVRS system was not offered because women were required to print, sign, and return the informed consent to the university coordinators, necessitating using the Internet system. Those who chose to participate by Internet were also offered the choice of responding to questionnaires every 2 or 4 weeks. Following study entry, women provided contact information, chose a personal identification number, and reminder method (text and/or email). Women could enroll at any time during their pregnancy, as long as it was prior to delivery, recognizing that not all women would be able to be followed throughout their entire pregnancies. Study data were collected using a study number and were maintained on a server that was separate from where contact details were stored.

Participants were recruited between October 1, 2012 (recruitment week 1) and January 31, 2014 (recruitment week 70) in DK, NL, and the UK; follow-up ended on March 28, 2014. Because of difficulties in arranging ethical approvals, the start date in PL was delayed until May 20, 2013 (recruitment week 34). Data were collected on medications used to treat illnesses including prescription and nonprescription medications, vaccines, x-rays, and various lifestyle factors, recreational drug use, and herbal products during current pregnancy or in the month preceding it; additionally, we collected basic demographics, education and ethnicity, and current and previous pregnancy history. Participants were also asked if they had used any medications that were not prescribed for them (borrowed medications) or had decided not to use medications that were prescribed. Follow-up questionnaires, provided at 2- or 4-week intervals as decided by the participant, asked about any changes to use of medications, and a final outcome questionnaire sought information about the birth outcome, including the presence of any birth defects. Participants were asked to tell us about medication use through a series of checkboxes of top 10 lists of medications that are commonly used during pregnancy specific to each country, as agreed by the study team. These lists were organized into medical conditions as a memory aid. Participants were also able to report other medication use as free texts.

In the UK, study participants were asked to provide their consent for linkage with primary care electronic health records (EHRs) in an effort to provide some validation of self-reported data, and in DK, consent for linkage with the national registries was mandatory for participation. Medications were classified according to their over-the-counter (OTC) or prescription status in DK, and applied to all other countries. Thus, medications were classified as available by prescription only (Rx only), available OTC, but may also be available by prescription in some countries depending on dose and various prescribing practices for pregnant women (OTC and Rx, or OTC or Rx), and OTC products not available by prescription (OTC only.) Those medications that were reported as being used during pregnancy and which can be obtained OTC and by prescription included acetylcysteine, acetylsalicylic acid, acyclovir, benzydamine, budesonide, fexofenedine, fluticasone, ibuprofen, lansoprazole, loperamide, pantoprazole, and paracetamol.

Results

Study Participants

Overall, 2521 women were enrolled in this study. Figure 1 shows a flowchart showing recruitment and retention through study close. There were 14 women (0.55%, 14/2521) who chose to provide information via IVRS and only 1 actually completed the baseline questionnaire, but did not provide information on birth outcome. Hence, results reported here are for those who participated via the Internet.

Internet Results

Of 43,068 people who clicked on the website, 23,536 (54.64%, 23,536/43,068) stayed for at least 30 seconds, and 2507 (10.65%) of 23,536 women enrolled and provided informed consent (DK 770; NL 568; PL 316; the UK 853). Some women discontinued participation before completing the baseline questionnaire, leaving 2065 who completed at least baseline data (DK 639; NL 476; PL 241; the UK 709). There were 43.00% (888/2065) of those who completed the baseline questionnaire that chose to respond every 2 weeks and 56.99% (1177/2065) every 4 weeks. The age distribution of participants in comparison to the general population of each country is shown in Table 1. Most participants reported themselves to be white (95.93%, 1981/2065 in all 4 countries). The educational status of participants, shown in Table 2, reveals that participants reported being highly educated, with 38.74% (800/2065) having completed some university or postgraduate education, and a particularly high rate of postgraduate education in PL. Most women had at least one previous pregnancy that resulted in a live birth, but 42.95% (887/2065) of participants were reporting their first pregnancies (38.3%, 245/639 in DK; 47.3%, 225/476 in NL; 50.6%, 122/241 in PL; and 41.6%, 295/709 in the UK); 23.00% (475/2065) of women enrolled were in their first trimester of pregnancy, 52.34% (1081/2065) in their second trimester, and 24.65% (509/2065) in their third trimester.

Figure 1. Recruitment and retention. WEB (Web, World Wide Web) is a method of accessing information over the medium of the Internet; IVRS: interactive voice response system; EDD: expected date of delivery.

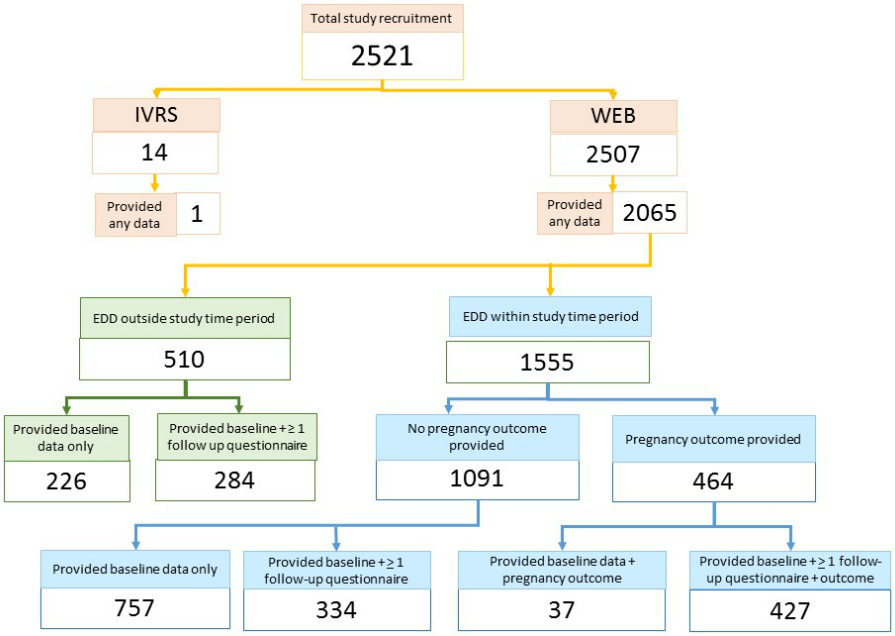


Table 1. Age of participants and referent population in each country.

			Age in years at end of pregnancy n (%)					
	Mean (SD)	n	<20	20-24	25-29	30-34	35-39	>40
Denmark [a]								
PROTECT[f]	31.5 (4.6)	639	1 (0.2)	43 (6.7)	214 (33.5)	237 (37.1)	118 (18.5)	26 (4.1)
National	30.9 (5.1)	55,225	750 (1.35)	6192 (11.21)	17,112 (30.98)	19,319 (34.98)	9769 (17.68)	2083 (3.77)
The Netherlands [b]								
PROTECT[f]	31.3 (4.3)	476	1 (0.2)	29 (6.1)	153 (32.1)	206 (43.3)	74 (15.5)	13 (2.7)
National 2012	30.9 (4.9)	173,085	2257 (1.30)	17,727 (10.24)	53,181 (30.72)	64,498 (37.26)	29,562 (17.07)	5860 (3.38)
Poland [c]								
PROTECT[f]	29.8 (4.2)	241	1 (0.4)	29 (12.0)	109 (45.2)	77 (32.0)	22 (9.1)	3 (1.2)
National	29.2 (NA[a])	370,932	14,522 (3.91)	63,158 (17.02)	131,373 (35.41)	110,192 (29.70)	43,554 (11.74)	8133 (2.19)
United Kingdom [d]								
PROTECT[f]	31.7 (5.1)	709	5 (0.7)	67 (9.4)	183 (25.8)	266 (37.5)	159 (22.4)	29 (4.1)
England & Wales 2013	30.0 (NA[e])	698,512	29,136 (4.17)	119,719 (17.13)	196,693 (28.15)	212,306 (30.39)	111,500 (15.96)	29,158 (4.17)

[a]The Birth Register at Statens Serum Institut; also see website [9].

[b]Stichting Perinatale Registratie Nederland: Perinatale Zorg in Nederland 2012. Utrecht: Stichting Perinatale Registratie Nederland 2013

[c]Central Statistical Office of Poland [10,11]

[d]UK Office for National Statistics, Age England & Wales 2013

[e]NA: not available

[f]PROTECT: Pharmacoepidemiological Research on Outcomes of Therapeutics by a European ConsorTium

Table 2. Educational levels of respondents compared with national statistics.

		Number (%) by highest education level				
	N	Legal part of school and 1st level of exams (age 16 years)	School and higher level exams	Some 3rd level education (university, etc)	University and some post graduate	Not stated
Denmark [a]						
PROTECT[e]	639	18 (2.8)	74 (11.7)	296 (46.7)	246 (38.8)	5 (0.8)
National	56,328	8639 (15.33)	18,488 (32.88)	17,431 (30.94)	8201 (14.55)	3569 (6.33)
The Netherlands [b]						
PROTECT[e]	476	22 (4.6)	110 (23.1)	161 (33.8)	183 (38.4)	0 (0)
National	2159	376 (17.41)	905 (41.91)	551 (25.52)	304 (14.08)	23 (1.06)
Poland [c]						
PROTECT[e]	241	4 (1.7)	21 (8.7)	56 (23.2)	158 (65.6)	2 (0.8)
National	17.0 million	4.2 million (24.4)	8.0 million (47.1)	0.6 million (3.7)	3.3 million (19.5)	0.9 million (5.3)
United Kingdom [d]						
PROTECT[e]	709	89 (12.6)	153 (21.6)	251 (35.4)	213 (30.0)	3 (0.4)
National	40.0 million	17.8 million (44.4)	8.7 million (21.8)	13.2 million (33)		0.3 million (0.8)

[a]Data from Statistics DK, Births 2012

[b]Data source: age: Stichting Perinatale Registratie Nederland. Perinatale Zorg in Nederland 2012. Utrecht: Stichting Perinatale Registratie Nederland, 2013; educational level: all women age 25-45, Statistics NL [12,13]

[c]The national figures for PL do not correspond completely with the educational levels used in PROTECT since level 4 for the national figures includes licentiate which is included in level 3 for PROTECT. Data source: Central Statistical Office of Poland Housing Census 2011 for women age 15 and over [10,14].

[d]The national figures for the UK are for women from 2011. The UK national figures for post graduate education not available.

[e]PROTECT: Pharmacoepidemiological Research on Outcomes of Therapeutics by a European ConsorTium

Medication Use During Pregnancy

There were 92.7% (1915/2065) of women that reported using at least one pregnancy-related medication including fertility medications, iron tablets, multivitamins, and folic acid. Excluding those pregnancy-related medications, 83.34% (1721/2065) of women reported using 1 or more medicines (range 1-16) during pregnancy or in the month preceding it. Figure 2 shows the distribution of number of medications by woman and country. Of all reported medications, 42.46% (2333/5494) were reported as being used “as needed” and 53.82% (2957/5494) were reported as being taken daily. Recognizing that some women reported more than one reason for their medication use, the most frequent reason for all reported medication use was for nervous system disorders (71.86%, 1484/2065), followed in descending order by 62.76% (1296/2065) for alimentary tract and metabolism, 38.01% (785/2065) for respiratory issues, 30.89% (638/2065) for genitourinary system or as sex hormones, and 25.66% (530/2065) as anti-infectives for systemic use.

The top 10 most frequently used medications and medication changes during pregnancy are shown in Table 3. There were 1230 (59.56%) of 2065 women who took at least one prescription medication (Rx) and may also have used OTC medications. By contrast, 23.58% (487/2065) women reported using medications generally available OTC, but never using any medications only available by prescription. These included 7.16% (148/2065) of our respondents, who reported that they chose not to take a medication that had been prescribed for them. The two most frequently reported indications for which a woman decided not to take a medication or decreased her dose (in consultation with her caregiver) were antidepressants and anti-inflammatories. Only 1.30% (27/2065) women reported using medicines that had been prescribed for a friend or family member, but not for them. The most frequent medications that were reported as having been shared were analgesics (oxycodone and paracetamol, n=4) and drugs for acid-related problems (n=8).

Table 3. The 10 most frequently reported medications used during pregnancy or during the month prior to pregnancy, and changes during pregnancy (rank ordered).

		Medications used during pregnancy or month prior to pregnancy		Rank order of top 10 medications				
				Medication		Changes during pregnancy		
		n/N	%	Rx	OTC	Not taken	Increased	Decreased[a]
Antibiotic								
	Pivmecillinam	67/2065	3.24	5				
Antidepressants								
	Citalopram	34/2065	1.65			2		5
	Fluoxetine	30/2065	1.45			7		7
	Sertraline	39/2065	1.89			10[b]		8
Antidiabetics								
	Insulin aspart	27/2065	1.30				5	
	Metformin	48/2065	2.32	9		10[b]		
Antifungal								
	Clotrimazole	135/2065	6.54		5			
Anti-inflammatories and pain								
	Acetylsalicyclic acid	48/2065	2.32				6	
	Diclofenac	16/2065	0.77			6		10[b]
	Ibuprofen	149/2065	7.21		4	1		1[b]
	Paracetamol	1483/2065	71.82		1	4	4	1[b]
	Paracetamol combination	80/2065	3.88		9			6
	Tramadol	28/2065	1.36			5		9
Anti-infectives								
	Amoxicillin	165/2065	8.00	1				
	Fluconazole	64/2065	3.10	6				
Antimigraine								
	Sumatriptan	31/2065	1.50			3		3[b]
Digestive disorders								
	Metoclopramide	47/2065	2.28	10				
	Alginic acid	291/2065	14.09		2		3	
	Ispaghula (psylla seeds)	90/2065	4.36		8			
	Lactulose	101/2065	4.89		6			
	Omeprazole	80/2065	3.88		10		9	
	Ordinary salt combinations (eg, calcium carbonate)	173/2065	8.37		3			
	Ranitidine	51/2065	2.46				7	
Ear, nose, and throat/decongestant								
	Xylometazoline	93/2065	4.50		7			
Reproductive								
	Clomifene	57/2065	2.76	8				

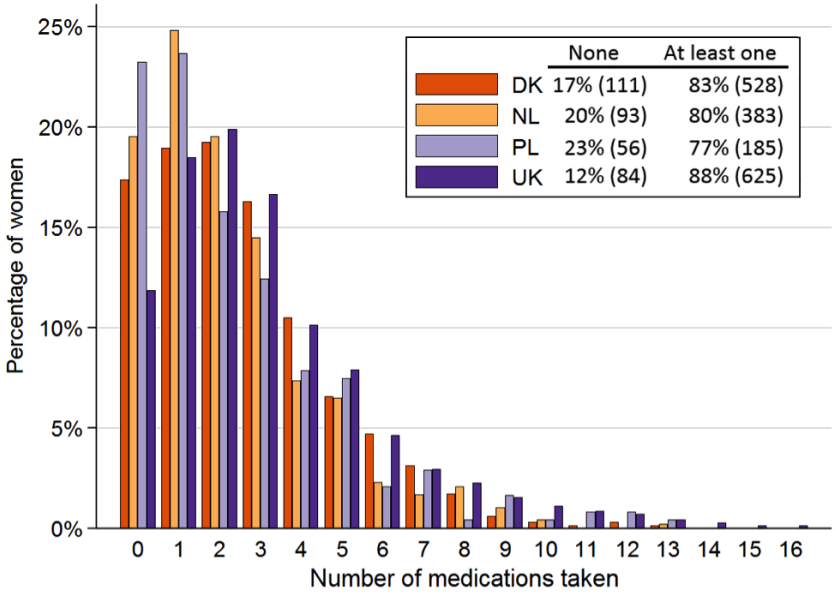
		Medications used during pregnancy or month prior to pregnancy		Rank order of top 10 medications				
				Medication		Changes during pregnancy		
		n/N	%	Rx	OTC	Not taken	Increased	Decreased[a]
	Ethinyl estradiol	5/2065	0.24			9		
	Hormonal contraceptives for systemic use	73/2065	3.54	4[c]		8		
	Progesterone	59/2065	2.86	7[c]				
Respiratory								
	Beclomethasone	30/2065	1.45				8	
	Salbutamol	96/2065	4.64	2			2	3[b]
	Terbutaline	32/2065	1.55				10	10[b]
Thyroid								
	Levothyroxine sodium	83/2065	4.02	3			1	

[a]Decreased, but not stopped

[b]Tied for place

[c]Probably represents use in the month before pregnancy

Figure 2. The number of different medications taken per woman by percentage in each country. Includes prescription and nonprescription medications; excludes herbals, fish oil, homeopathic, multivitamins, iron, vaccinations, and antimalarials. Denmark: DK, The Netherlands: NL, Poland: PL, United Kingdom: UK.



Use of Anesthetics, Cosmetic Surgery, and Alternative Medicines During Pregnancy

In addition to prescription medications, 7.31% (151/2065) of women reported using an anesthetic during pregnancy, with most (n=128) having received local anesthetics. Very few women (0.44%, 9/2065) reported having undergone a cosmetic procedure during pregnancy. Relatively few women used alternative medicines and dietary supplements during pregnancy (Table 4). The most frequently cited products were herbal (7.07%, 146/2065), followed by fish oil (4.60%, 95/2065).

Table 4. Percentage (number) women who used alternative medications and dietary supplements during pregnancy.

Medications	n/N	All (%)
Fish oil	95/2065	4.60
Homeopathic products	24/2065	1.16
Other dietary supplements	39/2065	1.89
Herbal products	146/2065	7.07
Neither	1761/2065	85.28
Total	2065/2065	100.00

Recreational Drugs During Pregnancy

Very few women (0.82%, 17/2065) reported using recreational drugs during pregnancy and among those recreational drug users, cannabis was by far the most widely used (65%, 11/17). The majority of women reported that they did not use any alcohol during pregnancy (76.90%, 1588/2065).

Retention and Data Quality

Study retention was relatively low with only 464 (29.84%) of 1555 women providing information on pregnancy outcome among those with due dates (plus 1 week) occurring while the study was active (Figure 1). Those who provided some follow-up were slightly more likely to continue in the study if they were taking 5 or more medications compared to those who dropped out after baseline (45.95%, 522/1136 vs 41.4%, 385/929). Further compared with those who continued in the study, women who discontinued participation after completing the baseline questionnaire were more likely to have been smokers at baseline (6.8%, 63/929 vs 2.46%, 28/1136), and to have volunteered to provide information monthly rather than every 2 weeks (64.7%, 601/929 vs 35.3%, 328/929), and were less likely to have completed university level coursework (71.8%, 667/929 vs 78.96%, 897/1136) or to be currently taking medications (3.5%, 33/929 vs .96%, 11/1136). There were no apparent differences between age, gravidity, or alcohol use for those who dropped out quickly compared with those who provided some follow-up.

In absence of source data verification, it is not possible to confirm the accuracy and reliability of all the data provided. Nevertheless, it was possible to compare self-reported data with available information in 2 countries, because all 639 participants in DK could be linked to their EHRs using the 10-digit civil registration number, and 79.0% (674/853) of women enrolled in the UK consented to EHR linkage. In the UK, the Health Improvement Network (THIN) database of pseudonymized primary care records was compared with the patient generated records. Because THIN is only a sample of the UK population, only 18 women were successfully linked to their EHRs in THIN and had data recorded during the period when the Pharmacoepidemiological Research on Outcomes of Therapeutics by a European ConsorTium (PROTECT) study subject was pregnant; 2 of those women were still pregnant at the close of PROTECT data collection.

We compared self-reported pregnancy outcome to the EHRs in the UK for 16/18 women who had reached their due date within the study period. Of the 16 whom we were able to match, 2 (12%) had an outcome (birth) noted in both PROTECT and their EHR. There were 8 (50%) of 16 women that had a pregnancy outcome recorded on EHR only, and 3 (19%) of 16 women reported only in PROTECT. In addition, information about the pregnancy outcome was not available in either PROTECT or the EHR for 3 of the 16 women (19%). It is possible that the THIN EHR may have been updated with information on their pregnancy outcomes after the data were provided for this study.

When data were compared from PROTECT with similar information for the women in the Danish Pharmacy register and from THIN in the UK (Table 5), some underreporting of medications was evident for women in PROTECT for chronic conditions, short-term conditions, and pregnancy-related medications. A small number of women also reported not taking prescribed medications, which may explain some differences between self-reports and database reports.

Table 5. Compatibility and discrepancy between PROTECT self-reported prescription use and electronic health care data in DK and the UK based on the whole follow-up period, $N_{PROTECT}/N_{DATABASE}$ (%).

	DK n=639		UK n=18
	All, n (%)	Excluding dispensed, but not taken[a], n (%)	All[b], n (%)
Drugs for chronic conditions	145/174 (83.3)	145/170 (85.3)	3/6 (50)
Drugs for occasional or short-term use	153/283 (54.1)	153/276 (55.4)	4/12 (33)
Pregnancy-related medications	78/118 (66.1)	78/114 (68.4)	6/10 (60)

[a]Adjusted for self-reported decision not to take a prescribed medication.

[b]None of these respondents in the UK reported a decision not to take a prescribed medication.

Discussion

Principal Findings

There is incomplete information about the efficacy and safety of treatments used by the broad array of patients seen in everyday clinical practice [15]. Could direct-to-patient research help fill that gap for pregnant women with information that could be used to investigate possible risk factors for negative birth outcomes in order to guide expectant mothers and their clinicians to make healthy pregnancy choices? Could direct-to-patient data also be used to understand treatment benefits and risks to inform decisions about personalized medicine [16]? This study shows that some pregnant women will volunteer to participate in pharmacovigilance-type studies and will report medications used in pregnancy, including both prescription and nonprescription medications, as well as other life style factors, like alcohol and drug use. A comparison of our respondents with demographics from each of their countries shows that our volunteers were broadly representative of the population density in each country, as well as age and parity, although they were more educated than their peers, and in some countries, were not as ethnically diverse as the underlying population (eg, 94.5%, 670/709 white in this study vs 86%, 48.2 million/56.1 million in the UK) [17]. To the extent that biologic responses to medication are not heavily dependent on education and ethnicity, these data would support the use of direct-to-patient research, as it may have benefits for understanding both safety and effectiveness. Further, it appears that interactive voice response is not a popular method for consumer responses. In this study, the IVRS was only chosen by 14 patients and of those, only 1 provided any usable data.

Recruitment and Retention

Loss to follow-up was high in this study, which without substantial improvement would lessen the value of direct-to-patient research for pharmacovigilance. This study, however, employed very minimal patient reminders, only unvarying reminder emails and/or text messages a few days before, and once after, a follow-up form was expected. No patient incentives were used and the Web study portal only provided study-related information and no other information to attract or maintain the interest of study participants. Using only these straight forward methods for retention, only 55.01% (1136/2065) of the subjects who enrolled in this study provided any follow-up, which may reflect, in part, the difficulty of filling out a detailed questionnaire and perhaps the unwanted focus on lifestyle factors that may negatively affect birth outcome. Only 29.83% (464/1555) of the women provided information about birth outcome. Among those who provided some follow-up, the loss-to-follow-up rates were more than doubled among women who enrolled in the second and third trimesters when compared to those who enrolled in the first trimester.

To further explore reasons that would affect recruitment and retention, we convened a small focus group of 45 first-time mothers in their first pregnancy from the DK, NL, PL, and the UK to assess their willingness and interest in participating in a study like PROTECT. Although most women reported that they would participate in a study like this purely for altruistic reasons, one of the most frequent comments was that some form of modest compensation would enhance the appeal of participation.

Prescription Reporting

Getting detailed, interpretable reports about medication use was challenging. We asked about medication use according to the indication for which it was being used, did not ask about dosage or route of administration, and did not distinguish between prescription and nonprescription medication because that varied from country to country and could change at any time during the study. Instead, we used a list of nonprescription medication for DK (the only country that was able to provide such a list) and applied that assumption to the other 3 countries, recognizing that in some countries, pregnant women receive prescriptions for medications that are available OTC in order to have those medications paid for by their insurance. We provided machine-prompts on the questionnaires for the top 10 medications in each country for each indication to guide data entry for medications, which was challenging to compile in 4 countries as proprietary names of the same products varied by country and by whether a prescription was required for a particular medication and dose. Nearly 25% (3136/12,699) of the medications used were reported as free text, rather than using the drop-down list available on the questionnaire for this purpose, which required substantial manual review and recoding. More consumer-friendly computerized methods of acquiring data about medication use would be helpful for future research.

Strengths of Direct-to-Patient Research

Despite the challenges we encountered in distinguishing medications that were obtained by prescription and those that were not, the extensive reporting of medication use here is a strong advantage for understanding the causal relations of drugs and adverse events. About 4 of 5 women (80%) reported using at least one medication during or immediately before pregnancy, a figure that is in line with some other estimates [18-22] and higher than others (eg, 50%) [23]. Our estimates may be higher than some other studies because our recruitment materials and informed consent documents described our interest in understanding medication use during pregnancy. Nonetheless, the information that our respondents provided corresponds fairly well with other reports. For example, a recent study of medications used by Medicaid recipients during pregnancy showed the most commonly dispensed medications during pregnancy were antibiotics and anti-infectives (nitrofurantoin, 21.6%; metronidazole, 19.4%; amoxicillin, 18.0%; azithromycin, 16.9%) and an antihistamine (promethazine, 13.5%) [23], whereas another US study showed the most commonly used prescription medication components during the first trimester of pregnancy were progestins from oral contraceptives, amoxicillin, progesterone, albuterol, promethazine, and estrogenic compounds [18]. In our study, the most common prescription medications were anti-infectives (amoxicillin) and antibiotics (pivmecillinam), as well as respiratory medication (salbutamol), thyroid medication (levothyroxine sodium), and hormonal medication (contraceptives). Interestingly, we noted that respondents might not always understand the indication for which they are taking prescribed medications. For example, in our validation study in the UK, physicians reported prescribing

no antidiabetic medications for the 18 women whose records were matched with PROTECT, yet 1 woman self-reported taking an antidiabetic medication.

Our estimate that 40.73% (841/2065) of pregnant women used nonprescription medicine is not much different from the limited published literature [19], even though the reference estimate may not include the identical nonprescription medications that were used in this study. In the US study of medications used during the first trimester of pregnancy, the most commonly used OTC medications included anti-inflammatories and pain medications (acetaminophen and ibuprofen) and medication for digestive disorders (docusate) [18], as did our study (paracetemol and ibuprofen; alginic acid and ordinary salt combinations like calcium carbonate). Of note, the methodology used here provided rich information about intermittent use of medication as well as nonprescription and complementary medication use, the details of which are rarely available from other sources. This pilot study gives some indication that women are willing to report their use of herbal medications, although the estimates from these 4 study countries (7.07%, 146/2065 of our respondents self-identified as using herbals during a pregnancy) are lower than the 6% and 11% reported in 2 US studies [20,21], and the 55% reported in Italy in 2009 [22].

Conclusions

Taken as a whole, it appears that direct to consumer (or patient) is a useful method for learning about use of prescription and nonprescription medication use, including medications that may be administered in hospitals, emergency rooms, or as outpatients, and in some cases, these patient-reported data are more complete than reliance on existing data like prescription registers and EHR [23]. Further, using the Internet to collect data directly from patients/consumers makes it easier to collect data regularly during a study. This steady data collection may reduce recall bias, which is particularly useful when investigating potential effects of medication use during pregnancy on birth outcomes, since such effects are largely dependent on gestational age at the time of exposure. In addition, because women could enter the study as soon as they became pregnant, it provides valuable information on exposure during the early weeks of pregnancy, which might not be available, or complete, using more traditional interview methods.

Could patient-reported information be a complete substitute for other types of pharmacovigilance? Although there appears to be some underreporting about medication use and perhaps herbals, self-reporting of prenatal exposures provides a more complete picture of factors which may contribute to an adverse outcome than does reliance purely on existing traditional methods of collecting data. The value of having information on medication use as well as other exposures that may not be readily available is appealing, for example, anesthesia, travel vaccinations, as well as other behaviors that patients may be reluctant to report directly to their physicians, such as use of recreational products like cocaine and marijuana. Such a rich data source would enhance our ability to evaluate individual teratogens as well as to look at various exposures, which may be risky when used in combination during pregnancy. Nonetheless, there are shortcomings to sole reliance on patient-reported data that make it difficult to use in the absence of supplementary data, including (1) accurate and complete product reporting of molecular entity, dose, and manufacturer; and (2) clinician-reported assessment of birth outcomes. Looking toward the future, perhaps the most effective method to assemble meaningful information about potential teratogens would be to combine patient-reported data with information on prescription and/or clinical validation from physicians or health care registers for major events and exposures of special interest.

Acknowledgments

The research leading to these results was conducted as part of the PROTECT consortium [24], which is a public-private partnership coordinated by the European Medicines Agency. The PROTECT project has received support from the Innovative Medicines Initiative Joint Undertaking under Grant Agreement No 115004, resources of which are composed of financial contribution from the European Union's Seventh Framework Programme (FP7/2007-2013) and EFPIA companies' in kind contribution. The views expressed are those of the authors only. In PL, the scientific work was cofinanced from the 2011-2014 allocation of funds for international projects from the Ministry of Science and Higher Education. Quintiles provided funding for one senior scientist (NAD). The authors gratefully acknowledge the help of A Latos-Bieleńska, Department of Medical Genetics, Poznan University of Medical Sciences; J-P Balling, Lundbeck; L Comisky, Genzyme; Rebecca Groves, International Alliance of Patients' Organizations; O de Mol, Genzyme; B Patel, Amgen; S Stevens, Institute of Cellular Medicine, Newcastle University.

Conflicts of Interest

None declared.

References

1. Herbst AL, Ulfelder H, Poskanzer D. Adenocarcinoma of the vagina. Association of maternal stilbestrol therapy with tumor appearance in young women. N Engl J Med 1971 Apr 15;284(15):878-881. [doi: 10.1056/NEJM197104222841604] [Medline: 5549830]
2. Clemow DB, Nolan M, Michaels D, Kogelnik A, Cantrell SA, Dewulf L. Medicines in pregnancy forum: Proceedings on ethical and legal considerations. Therapeutic Innovation & Regulatory Science 2015 Feb 27;49(3):326-332. [doi: 10.1177/2168479015572373]

3. Brody T, Larner J, Minneman K. Perinatal/neonatal pharmacology. In: Human pharmacology: Molecular to clinical. St. Louis: Mosby; 1998.
4. Mitchell AA. Systematic identification of drugs that cause birth defects--a new opportunity. N Engl J Med 2003 Dec 25;349(26):2556-2559. [doi: 10.1056/NEJMsb031395] [Medline: 14695418]
5. Margulis AV, Mittleman MA, Glynn RJ, Holmes LB, Hernández-Díaz S. Effects of gestational age at enrollment in pregnancy exposure registries. Pharmacoepidemiol Drug Saf 2015 Apr;24(4):343-352. [doi: 10.1002/pds.3731] [Medline: 25702683]
6. Jick H. Recall "error" in interview studies of past drug use [Letter]. Am J Public Health 1982 Apr;72:405 [FREE Full text]
7. Rhodes SD, Diclemente R, Yee LJ, Hergenrather K. Factors associated with testing for hepatitis C in an internet-recruited sample of men who have sex with men. Sex Transm Dis 2001 Sep;28(9):515-520. [Medline: 11518868]
8. Dreyer NA, Blackburn S, Hliva V, Mt-Isa S, Richardson J, Jamry-Dziurla A, et al. Balancing the interests of patient data protection and medication safety monitoring in a public-private partnership. JMIR Med Inform 2015;3(2):e18 [FREE Full text] [doi: 10.2196/medinform.3937] [Medline: 25881627]
9. Sundhud. URL: http://www.esundhed.dk/sundhedsregistre/MFR/Sider/MFR01.aspx [accessed 2015-11-22] [WebCite Cache ID 6dETM2KVO]
10. GUS. URL: http://stat.gov.pl/ [accessed 2015-11-22] [WebCite Cache ID 6dETpmMj3]
11. Demographfia. URL: http://demografia.stat.gov.pl/bazademografia [accessed 2015-11-21] [WebCite Cache ID 6dEU6JTyw]
12. CBS. URL: http://www.cbs.nl/nl-NL/menu/home/default.htm [WebCite Cache ID 6dEUMa4fA]
13. statline. URL: http://statline.cbs.nl/Statweb/publication/?DM=SLNL&PA=71822NED&D1=0&D2=0,l&D3=2-3&D4=0-4&D5=a&D6=0&D7=l&HDR=T,G2,G1,G5,G6&STB=G4,G3&VW=T [WebCite Cache ID 6dEUZTsIn]
14. BAZA Demografia. URL: http://demografia.stat.gov.pl/bazademografia/Tables.aspx [accessed 2015-11-22] [WebCite Cache ID 6dEUqTAAc]
15. Chandra A, Khullar D, Lee T. Addressing the challenge of gray-zone medicine. N Engl J Med 2015 Jan 15;372(3):203-205. [doi: 10.1056/NEJMp1409696] [Medline: 25587945]
16. Jameson JL, Longo D. Precision medicine--personalized, problematic, and promising. N Engl J Med 2015 Jun 4;372(23):2229-2234. [doi: 10.1056/NEJMsb1503104] [Medline: 26014593]
17. Office for National Statistics. Ethnicity and national identity in England and Wales 2011. London, UK: Office for National Statistics; 2012 Dec 11. URL: http://www.ons.gov.uk/ons/rel/census/2011-census/key-statistics-for-local-authorities-in-england-and-wales/rpt-ethnicity.html [accessed 2015-07-14] [WebCite Cache ID 6a1SgjWAj]
18. Thorpe PG, Gilboa S, Hernandez-Diaz S, Lind J, Cragan J, Briggs G, National Birth Defects Prevention Study. Medications in the first trimester of pregnancy: Most common exposures and critical gaps in understanding fetal risk. Pharmacoepidemiol Drug Saf 2013 Sep;22(9):1013-1018 [FREE Full text] [doi: 10.1002/pds.3495] [Medline: 23893932]
19. Werler M, Mitchell A, Hernandez-Diaz S, Honein M. Use of over-the-counter medications during pregnancy. Am J Obstet Gynecol 2005 Sep;193(3 Pt 1):771-777. [doi: 10.1016/j.ajog.2005.02.100] [Medline: 16150273]
20. Louik C, Gardiner P, Kelley K, Mitchell A. Use of herbal treatments in pregnancy. Am J Obstet Gynecol 2010 May;202(5):439.e1-439.e10 [FREE Full text] [doi: 10.1016/j.ajog.2010.01.055] [Medline: 20452484]
21. Broussard C, Louik C, Honein M, Mitchell A, National Birth Defects Prevention Study. Herbal use before and during pregnancy. Am J Obstet Gynecol 2010 May;202(5):443.e1-443.e6. [doi: 10.1016/j.ajog.2009.10.865] [Medline: 20035911]
22. Cuzzolin L, Francini-Pesenti F, Verlato G, Joppi M, Baldelli P, Benoni G. Use of herbal products among 392 Italian pregnant women: Focus on pregnancy outcome. Pharmacoepidemiol Drug Saf 2010 Nov;19(11):1151-1158. [doi: 10.1002/pds.2040] [Medline: 20872924]
23. Haerskjold A, Henriksen L, Way S, Malham M, Hallas J, Pedersen L, et al. The Danish National Prescription Registry in studies of a biological pharmaceutical: Palivizumab - validation against two external data sources. Clin Epidemiol 2015;7:305-312 [FREE Full text] [doi: 10.2147/CLEP.S73355] [Medline: 26056490]
24. Protect. URL: http://www.imi-protect.eu/ [accessed 2015-11-22] [WebCite Cache ID 6dEhc77cy]

Abbreviations

DK: Denmark
EHRs: electronic health records
IVRS: interactive voice response system
NL: the Netherlands
OTC: over-the-counter
PL: Poland
PROTECT: Pharmacoepidemiological Research on Outcomes of Therapeutics by a European ConsorTium
UK: United Kingdom

Edited by G Eysenbach; submitted 14.07.15; peer-reviewed by C Broussard; comments to author 05.08.15; revised version received 21.08.15; accepted 20.09.15; published 22.12.15.

Please cite as:
Dreyer NA, Blackburn SCF, Mt-Isa S, Richardson JL, Thomas S, Laursen M, Zetstra-van der Woude P, Jamry-Dziurla A, Hliva V, Bourke A, de Jong-van den Berg L
Direct-to-Patient Research: Piloting a New Approach to Understanding Drug Safety During Pregnancy
JMIR Public Health Surveill 2015;1(2):e22
URL: http://publichealth.jmir.org/2015/2/e22/
doi:10.2196/publichealth.4939
PMID:27227140

Original Paper

Tracking Hookah Bars in New York: Utilizing Yelp as a Powerful Public Health Tool

Philip B Cawkwell[1], BA; Lily Lee[1,2]; Michael Weitzman[1,3,4,5], MD; Scott E Sherman[3,5,6,7], MD, MPH

[1]Department of Pediatrics, New York University School of Medicine, New York, NY, United States

[2]Brooklyn College, Brooklyn, NY, United States

[3]College of Global Public Health, New York University, New York, NY, United States

[4]Department of Environmental Health, New York University School of Medicine, New York, NY, United States

[5]NYU/Abu Dhabi Public Health Research Center, Abu Dhabi, United Arab Emirates

[6]Department of Population Health, New York University School of Medicine, New York, NY, United States

[7]Department of Medicine, VA New York Harbor Healthcare System, New York, NY, United States

Corresponding Author:
Philip B Cawkwell, BA
Department of Pediatrics
New York University School of Medicine
550 1st Avenue
New York, NY 10016
United States
Phone: 1 212 263 7300
Fax: 1 212 263 7300
Email: Philip.Cawkwell@med.nyu.edu

Abstract

Background: While cigarette use has seen a steady decline in recent years, hookah (water pipe) use has rapidly increased in popularity. While anecdotal reports have noted a rise in hookah bars, methodological difficulties have prevented researchers from drawing definitive conclusions about the number of hookah bars in any given location. There is no publicly available database that has been shown to reliably provide this information. It is now possible to analyze Internet trends as a measure of population behavior and health-related phenomena.

Objective: The objective of the study was to investigate whether Yelp can be used to accurately identify the number of hookah bars in New York State, assess the distribution and characteristics of hookah bars, and monitor temporal trends in their presence.

Methods: Data were obtained from Yelp that captures a variety of parameters for every business listed in their database as of October 28, 2014, that was tagged as a "hookah bar" and operating in New York State. Two algebraic models were created: one estimated the date of opening of a hookah bar based on the first Yelp review received and the other estimated whether the bar was open or closed based on the date of the most recent Yelp review. These findings were then compared with empirical data obtained by Internet searches.

Results: From 2014 onward, the date of the first Yelp review predicts the opening date of new hookah bars to within 1 month. Yelp data allow the estimate of such venues and demonstrate that new bars are not randomly distributed, but instead are clustered near colleges and in specific racial/ethnic neighborhoods. New York has seen substantially more new hookah bars in 2012-2014 compared with the number that existed prior to 2009.

Conclusions: Yelp is a powerful public health tool that allows for the investigation of various trends and characteristics of hookah bars. New York is experiencing tremendous growth in hookah bars, a worrying phenomenon that necessitates further investigation.

(JMIR Public Health Surveill 2015;1(2):e19) doi:10.2196/publichealth.4809

KEYWORDS

hookah; hookah bar; Internet; public health; Yelp

Introduction

Hookah Smoking Perceived Less Dangerous Than Cigarettes

In recent years, cigarette use has had a dramatic steady decline [1-3], whereas hookah (water pipe) use has rapidly become more popular. National Adult Tobacco Survey data show that 9.8% of adults in the United States report having ever used hookah and 1.5% currently use this alternative tobacco product [4]. Among US high-school seniors, 18% report having used hookah in the past year [5]. There is growing evidence that hookah smoking may be at least as harmful as cigarette smoking [6-8]. Research has shown that the tobacco specifically used for hookah—shisha—delivers tar, nicotine, and carbon monoxide in higher doses than cigarette use [9-11]. Hookah use has also been linked to lung cancer [12], decreased pulmonary and cardiovascular function [13,14], infertility [15], and low birth weight [12]. Furthermore, secondhand hookah smoke contains the same constituents resulting from tobacco combustion, as well as additional chemicals and carcinogens from the charcoal used to heat the product [16,17], thus posing serious health risks not only to the user, but also to hookah bar employees and nonsmoking patrons [7,17,18]. Unfortunately, numerous studies worldwide consistently report that hookah smoking is perceived to be safer and less addictive than cigarettes by both public and health care providers [8,9,19-23].

In the last decade, Internet use in the United States has expanded rapidly as millions of consumers have new ways to access the Internet, from laptops to mobile phone to tablets and other devices [24]. The Internet is used by as many as 93% of high-school students and young adults [25]. The Internet also has emerged as a new resource for tracking public health trends [26]. It is now possible to analyze Internet trends as a measure of population behavior and health-related phenomena. The social media platform Twitter has been analyzed to show how users perceive and respond to emerging tobacco products [27,28]. One database, Google Trends, a publicly available resource, allows users to compare Internet search frequency over time in various regions of the world [29,30]. Researchers have used data drawn from this source to create models that accurately predict chronic disease risk factors such as alcohol consumption by measuring Web search activity [31]. This mechanism has also been used to demonstrate the increasing popularity of hookah smoking in the United States as measured by increased search volume for hookah and related terms [32].

Social Media and Yelp

In recent years, new forms of social media and networking have also risen. Yelp is one such site in which patrons review and share information about businesses and their services, thereby allowing local businesses to publicize themselves to the public [33,34]. Based out of San Francisco, the company has set up online communities in almost every major city in the United States and worldwide. In 2014, it received an average of 135 million visitors per month and more than 67 million total reviews [35]. Yelp is ranked in the top 40 in the United States and top 150 worldwide in number of daily visitors to the website [36].

To date, no study has examined the number, distribution, and potential proliferation of hookah bars. There is no publically available database of hookah bars at either a state or federal level. Yelp has an unharnessed, but immense, potential for providing critical public health data. In this paper, we report whether data from Yelp can be used to accurately answer the following three questions: (1) What is the number of hookah bars throughout a specific geographic region and how accurately can Yelp identify the hookah bars?; (2) What are the distribution and characteristics of hookah bars (such as whether they serve alcohol or have live music)?; and (3) What are the temporal trends in hookah bars?

Methods

Data

Data were provided by Yelp in the form of a spreadsheet file that captured a variety of parameters for every business listed in their database as of October 28, 2014 that was tagged as a “hookah bar” and operating in the state of New York. We also obtained similar data for “wine bar” to serve as a contemporaneous control. These parameters included the business address, categorization within the Yelp database (eg, “bars,” “hookah bars,” “lounges”), business hours, and attributes such as whether alcohol is served, the ambiance, whether Wi-Fi is available, etc. In addition, every review written for each of these businesses was provided in a separate file with the reviewer’s username removed. According to Yelp’s policy, business listings are not removed after the location is determined to be closed. These businesses may feature less prominently in search results, but this did not affect the database file that we received.

Analyses

Two simple algebraic models were created from the data provided by Yelp. The first was created to estimate the date of opening of a hookah bar based on the date of its first Yelp review. To do this, we attempted to determine the actual date of opening for each hookah bar via Internet searches. A precise opening date could be established for 46 hookah bars. Then, the actual date of opening was compared with the date of the first published Yelp review for each hookah bar. The algebraic model compared these differences on a yearly basis and used the average difference between verified date of opening and first Yelp review as the predicted date of opening for each hookah bar. Because of low review frequency in the early years of Yelp, it was not practical or accurate to use the date of the first published review as a predictor of true opening date prior to 2010.

A second algebraic model was created that predicted whether or not a hookah bar had shut down based on the date of its latest Yelp review. The Yelp website allows its users to mark locations as closed, and this information is contained within the Yelp database. We verified that all 24 of these locations were indeed closed with phone calls to the location and Internet searches. We then combed the remaining database to find whether there were other closed locations that had not been marked as such, and 7 were identified with further phone calls and Internet searches. Of the 24 that were marked closed on Yelp, 17 had

not had a Yelp review written in over 1 year. Next, we computed the number of months since the last Yelp review for each hookah bar. The range was 0-65 months, the mean was 6.3 months, the median was 1 month, and the mode was 0 months (49 bars). The model considered a hookah bar as closed if it had not received a review in the past 6 months.

To determine the geographic location of hookah bars, the addresses of all hookah bars in the Yelp database were entered into Tableau Public, a freeware data visualization software program. This allowed for the generation of a heat map that examined the density of hookah bars by zip code. We present only the heat map for New York City here as it captures 121/137 (88.3%) of the hookah bars in New York State.

The reviews received by hookah bars in New York were examined in a summative fashion by examining the total number of reviews received each month from the date of the first review to October 2014. This was compared with the number of reviews received by all of the wine bars in New York, which served as a control.

Results

Opening Dates of Hookah Bars in New York State

The Yelp database identified 137 hookah bars in the state of New York. Using a variety of aforementioned search methods, the opening date was identified for 46 currently open hookah bars. A total of 43 of the 46 opened after Yelp's founding in 2004 and thus could be used for comparison. For bars that opened in 2010 or later, the date of the first Yelp review was a reliable proxy measure for the date of the bar opening to within 6 months (Table 1). For reviews written after January 1, 2014, only a 1-month correction factor was necessary. The date of the first Yelp review for each hookah bar was identified and this information was used to create an algebraic model that predicts the date of opening. The model predicts that 83 of the 113 (73.5%) currently open hookah bars were founded in 2009 or later, whereas the empirical data found that 40 of 43 (93%) bars opened in 2009 or later (Figure 1).

Table 1. Length of time between hookah bar opening and first Yelp review by year.

Real year opened	Number of hookah bars	Difference between date opening and first Yelp review (months)[a]
2006	2	34.5
2007	0	—
2008	1	22
2009	2	13
2010	6	3.5
2011	3	5.33
2012	11	3.18
2013	6	6.33
2014	12	0.75

[a]Difference calculated for the 43 hookah bars where date of opening could be verified and where the date was after Yelp's founding (ie, after October 2004).

Identifying Closed Hookah Bars

The Yelp database identified 24 hookah bars as closed. An algebraic model was created, using the date of the most recent Yelp review for every hookah bar in the database, to predict whether a hookah bar was open or closed. Any hookah bar that did not have a Yelp review within the last 6 months was predicted to be "closed." A comparison of the model with the Yelp database data is provided in Tables 2 and 3. The specificity of the model is 94.7% (compared with 100% for the Yelp database "closed" tool), whereas the sensitivity of the model is 90.3% (compared with 77.4% for the raw Yelp data); the model identifies more closed bars, but makes more errors by falsely stating that bars are closed when they are open.

Figure 1. Distribution of new hookah bars by year. Actual data represents the 43 hookah bars for which a date of opening could be identified. The algebraic model uses the date of the first Yelp review for each of the 137 bars in NY (New York) to approximate its date of opening.

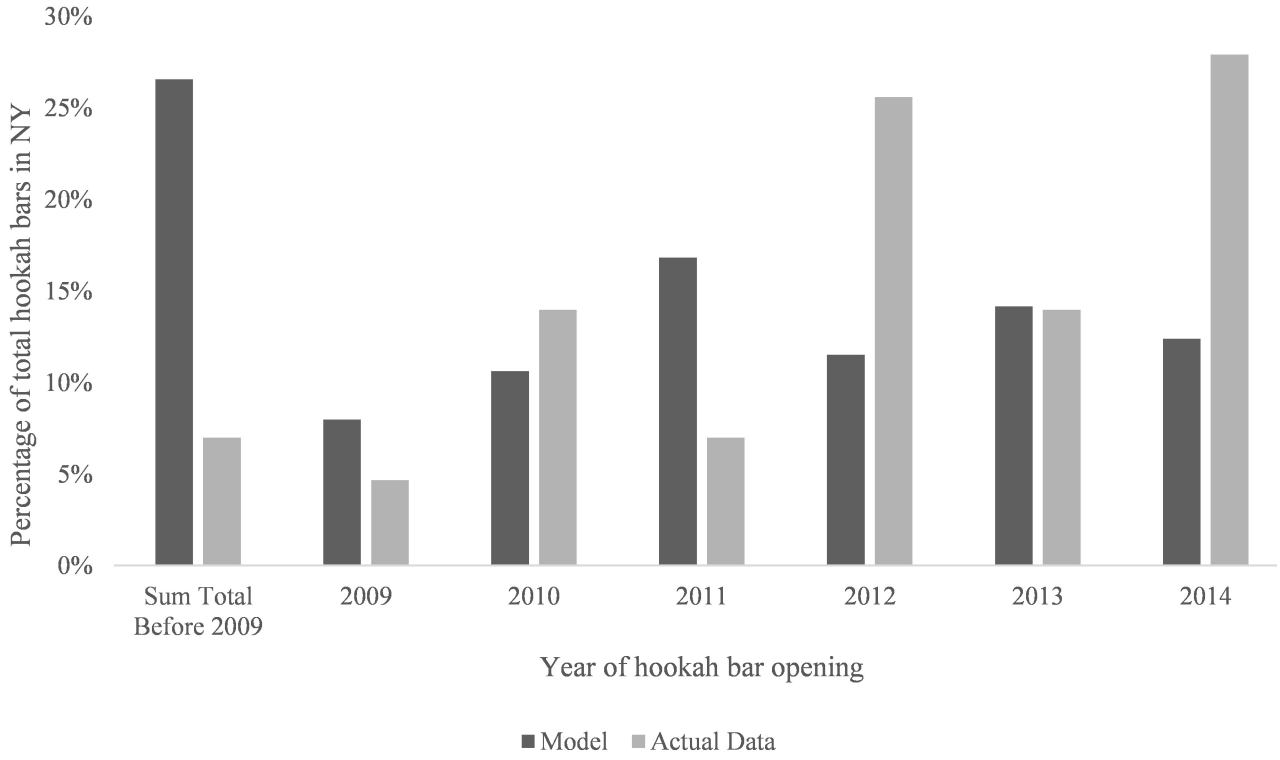


Table 2. Contingency table: Yelp-determined open/closed status of hookah bars.

	Yelp marked closed[a]	Yelp marked open
Closed hookah bars[b]	24	7
Open hookah bars	0	113

[a]"Yelp marked closed/open" refers to whether Yelp has marked the hookah bar as open or closed in their database.

[b]The closed/open hookah bar cells reflect the true status based on telephone calls and Internet searches.

Table 3. Contingency table: Model determined open/closed status of hookah bars.

	Model marked closed[a]	Model marked open
Closed hookah bars	28	3
Open hookah bars	6	107

[a]The algebraic model uses the date of the last review for the hookah bar to estimate whether it is open or closed. Bars that have not received a review within the last 6 months are estimated to be closed.

Hookah Bar Locations

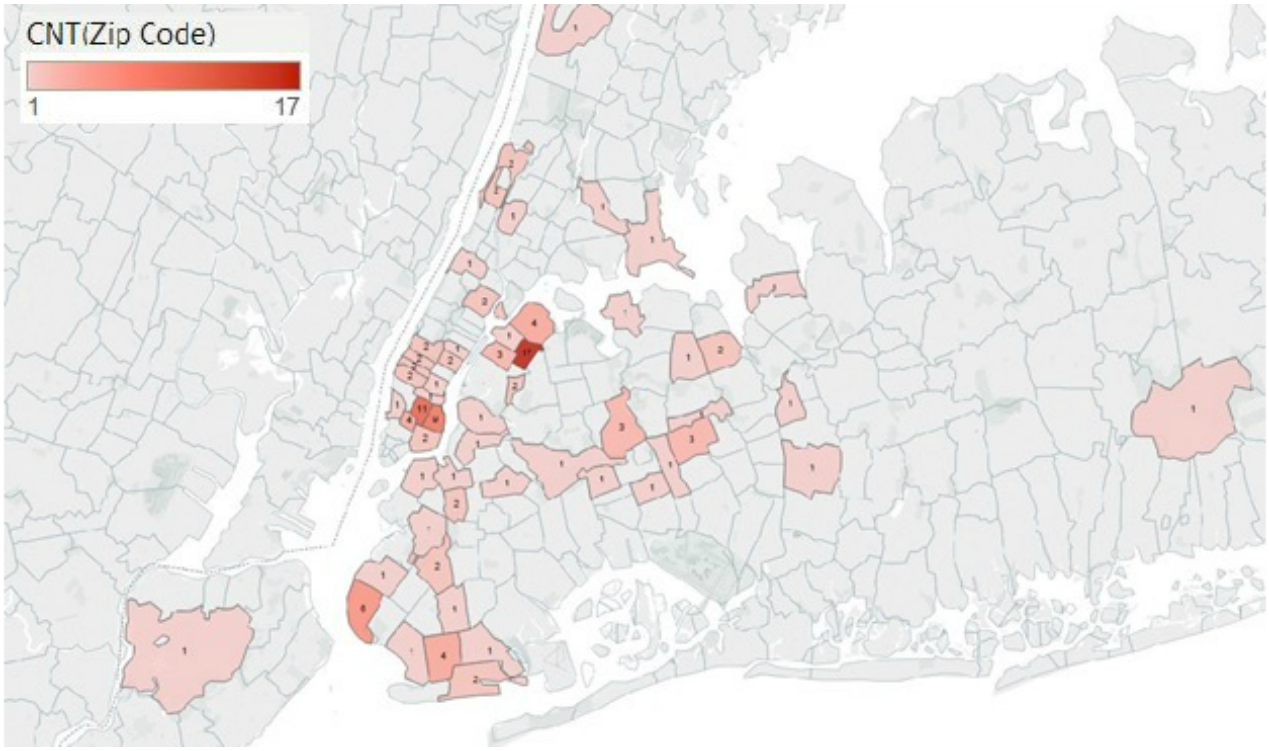
The location and distribution of the hookah bars were also examined. An overwhelming majority (121/137) of the hookah bars in the state of New York are located in one of New York City's 5 boroughs. The majority are clustered in Queens (n=43), Manhattan (n=43), and Brooklyn (n=26). Figure 2 shows a heat map with the locations of hookah bars by zip code. The greatest cluster occurs in zip code 11103, a subdivision of Astoria, Queens, a 0.71-mi^2 area [37] with 17 hookah bars. The zip codes 10002, 10003, and 10009 comprise Manhattan's Lower East Side and 10001-10006 comprise Astoria, Queens, according to the United Hospital Fund classification [38]. These are the 2 neighborhoods in New York with the highest density of hookah bars.

Unique characteristics of the hookah bars were examined (see Multimedia Appendix 1). The information provided ranges from whether the bars accept credit cards (115/137, 83.9% yes) to whether patrons describe the ambience as "trendy" (20/137, 14.6% yes).

Finally, temporal trends were evaluated by analyzing the individual reviews received by each hookah bar. The first review of a New York hookah bar occurred on August 20, 2005; the number of hookah bar reviews has grown manifold since then. From August 2005 to August 2008, there were an average of 7.2 reviews of hookah bars per month. From August 2013 to 2014, this number jumped to 118.8. We compared the cumulative number of reviews for hookah bars with those for wine bars to control for the growth rate of Yelp. From October

2008 to 2014, the total number of hookah bar reviews increased 17-fold, compared to a 10-fold increase for wine bars.

Figure 2. Heat map of location of hookah bars in New York City by zip code.



Discussion

Principal Findings

In this study, we used a dataset provided by Yelp to create a simple algebraic model that identifies when a hookah bar was founded based on the date of its first review. The model estimates that there have been more new hookah bars from 2012 to 2014 than the number that existed prior to 2009. While others have anecdotally noted the evidence of rapid expansion of hookah bars [39], we have taken this a step further by identifying the year-by-year expansion seen in New York. This suggests that retailers in the form of new hookah bars have met the increase in demand for hookah. Salloum et al [32] used the novel mechanism of Internet search queries to demonstrate the expanding popularity of hookah use. The rapidly increasing frequency of reviews for hookah bars relative to wine bars on Yelp shown here seems to corroborate this. Others have conducted more traditional surveys and analyses that agree with the notion that there is high demand for hookah [5,40,41].

We also used the Yelp data to map the locations of hookah bars in New York, and specifically focused on the distinct clusters in New York City because this was where the vast majority of hookah bars were located. One cluster was Astoria, Queens, which contained 27 hookah bars, and another was the Lower East Side, which contained 22. This finding is thought provoking because a large population of college students reside in this area and a large population of immigrants [42], especially Arabs, reside in Astoria [43]. This is significant because these populations typically have high rates of tobacco use [44,45]. Others have corroborated the idea that hookah bar expansion tends to occur near colleges [46] and to attract adolescents [47]. It is not entirely surprising that business owners would want to set up shop in areas with high demand.

Limitations

There are several limitations to this study. The first is that the precision of our algebraic models was limited by the temporally skewed data. The earliest verified date of opening of a hookah bar in New York City in our database is 1977 (the hookah component was not added until the 2000s); Yelp, however, was founded in October 2004, and this specific bar did not appear until January 2006. Thus, the model could only say that a bar was likely to have been founded “pre-2006.” It is inherently more accurate with newer bars and there is a higher level of uncertainty about the opening date of bars founded before 2009, for which the model could not accurately identify the opening date to within a year. In addition, 38 of the 46 hookah bars that we could identify had an official opening date that occurred in 2010 or later. It is possible that there is some form of recency effect occurring, as we are more likely to be able to identify more newly opened hookah bars that may have promoted themselves via the Internet. This is reflected in Figure 1, as few bars were identified with an opening date prior to 2009 and yet the model predicted many bars to have opened in that range based on their first date of Yelp review. Yelp’s expanding popularity ensured that a hookah bar was likely to have a review posted on Yelp within the first 6 months of its opening from 2010 to 2013, and within a month from 2014 onward, which results in higher confidence with the more recent data. Our comparison of hookah bar review frequency with wine bar frequency is a relative control, as it is unclear whether they have been increasing or decreasing in popularity. Thus, this comparison can only serve to provide some context for the rate of increase in hookah bar reviews. Next, our algebraic model

was intentionally designed to be simplistic. This allows for a robust and rapid estimation of whether a hookah bar is open or closed and when it was founded based simply on the date of its first and most recent reviews. As a drawback, however, it is not as precise as could potentially be achieved with more comprehensive measures.

Yelp is a powerful tool and is getting still more popular; we noted a 20-fold increase in monthly reviews for hookah bars between January 2007 and 2014. In spite of this, there have been very few public health studies that utilize its vast database, and only one that used it in the context of hookah. Primack et al [48] utilized Yelp to locate hookah bars in municipalities with clean air laws. Sussman et al [34] coded Yelp reviews in Los Angeles, California, for Vape shops to analyze consumer beliefs and behaviors. Other medical studies that use Yelp data focus on analyzing reviews of physicians who are listed on Yelp [49-51] and on food safety [52,53].

Yelp offers both public health researchers and potentially local health departments the ability to quickly and accurately spot trends. We found that on average, beginning in 2014, a new hookah bar will appear on Yelp within 1 month of its opening. This allows for rapid identification of an expanding market. This could be particularly useful for local health departments. In New York City, for example, the health department is very active in enforcing health standards; they recently undertook an operation to identify which hookah bars were serving shisha-containing tobacco, which is a violation of the 2002 Smoke-Free Air Act [54]. Utilization of Yelp data would be beneficial because it would allow health departments to rapidly identify new bars. The same would apply to researchers seeking to measure the effects on working in a hookah bar or track their expanding popularity.

Some data contained within the Yelp database are more immediately useful than others are. It is readily apparent how knowledge of location of the hookah bars could be useful to researchers, health departments, and policy makers. However, there is a trove of other data collected for every hookah bar which includes whether they have parking, television, live music, offer outdoor seating, accept credit cards, or if they serve alcohol. Whether there are important findings or implications contained in these data remain to be seen, but it is clear that Yelp is an important, but underutilized, public health tool.

Conclusions

In summary, the findings presented in this paper corroborate that data from Yelp do in fact accurately identify the supply of hookah bars and their distribution and characteristics, as well as allowing for monitoring changes in their presence over time. In the state of New York, these data demonstrate both an increasing number of such venues and substantial geographic clustering. As emerging health epidemics like hookah use grow, public health officials and researchers would be well served to consider innovative sources such as Yelp for data analysis.

Acknowledgments

This work was supported in part by the NYU/Abu Dhabi Public Health Research Center (Drs Sherman and Weitzman). Dr Sherman is also supported by a grant from the National Institute on Drug Abuse (Grant No 1K24DA038345) and Dr Weitzman also is supported by a grant from the National Institute of Environmental Health (Grant No NIH/NCI 3 P30 CA016087-33S1). The content of this work is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health or the US Government.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Attributes of the 137 hookah bars in New York, as described by Yelp.

[PDF File (Adobe PDF File), 29KB - publichealth_v1i2e19_app1.pdf]

References

1. Centers for Disease Control and Prevention (CDC). Tobacco product use among middle and high school students—United States, 2011 and 2012. MMWR Morb Mortal Wkly Rep 2013 Nov 15;62(45):893-897 [FREE Full text] [Medline: 24226625]
2. Johnston LD, O'Malley PM, Miech RA, Bachman JG, Schulenberg JE. Monitoring the Future National Results on Drug Use: 1975-2013: Overview, Key Findings on Adolescent Drug Use. Ann Arbor, MI: The University of Michigan; 2014. URL: http://www.monitoringthefuture.org/pubs/monographs/mtf-overview2013.pdf [accessed 2015-06-09] [WebCite Cache ID 6ZA9VWdLi]
3. Substance Abuse and Mental Health Services Administration. Results from the 2012 National Survey on Drug Use and Health: Summary of National Findings, NSDUH Series H-46, HHS Publication No (SMA) 13-4795. Rockville, MD: Substance Abuse and Mental Health Services Administration; 2013. URL: http://www.samhsa.gov/data/sites/default/files/NSDUHresults2012/NSDUHresults2012.pdf [accessed 2015-11-17] [WebCite Cache ID 6ZA9c1qs6]
4. Salloum RG, Thrasher JF, Kates FR, Maziak W. Water pipe tobacco smoking in the United States: Findings from the National Adult Tobacco Survey. Prev Med 2015 Feb;71:88-93. [doi: 10.1016/j.ypmed.2014.12.012] [Medline: 25535678]

5. Palamar JJ, Zhou S, Sherman S, Weitzman M. Hookah use among U.S. high school seniors. Pediatrics 2014 Aug;134(2):227-234 [FREE Full text] [doi: 10.1542/peds.2014-0538] [Medline: 25002664]
6. Cobb C, Ward KD, Maziak W, Shihadeh AL, Eissenberg T. Waterpipe tobacco smoking: An emerging health crisis in the United States. Am J Health Behav 2010;34(3):275-285 [FREE Full text] [Medline: 20001185]
7. Daher N, Saleh R, Jaroudi E, Sheheitli H, Badr T, Sepetdjian E, et al. Comparison of carcinogen, carbon monoxide, and ultrafine particle emissions from narghile waterpipe and cigarette smoking: Sidestream smoke measurements and assessment of second-hand smoke emission factors. Atmos Environ (1994) 2010 Jan 1;44(1):8-14 [FREE Full text] [doi: 10.1016/j.atmosenv.2009.10.004] [Medline: 20161525]
8. Noonan D, Kulbok PA. New tobacco trends: Waterpipe (hookah) smoking and implications for healthcare providers. J Am Acad Nurse Pract 2009 May;21(5):258-260. [doi: 10.1111/j.1745-7599.2009.00402.x] [Medline: 19432909]
9. Eissenberg T, Shihadeh A. Waterpipe tobacco and cigarette smoking: Direct comparison of toxicant exposure. Am J Prev Med 2009 Dec;37(6):518-523 [FREE Full text] [doi: 10.1016/j.amepre.2009.07.014] [Medline: 19944918]
10. Knishkowy B, Amitai Y. Water-pipe (narghile) smoking: An emerging health risk behavior. Pediatrics 2005 Jul;116(1):e113-e119. [doi: 10.1542/peds.2004-2173] [Medline: 15995011]
11. Shihadeh A. Investigation of mainstream smoke aerosol of the argileh water pipe. Food Chem Toxicol 2003 Jan;41(1):143-152. [Medline: 12453738]
12. Akl EA, Gaddam S, Gunukula SK, Honeine R, Jaoude PA, Irani J. The effects of waterpipe tobacco smoking on health outcomes: A systematic review. Int J Epidemiol 2010 Jun;39(3):834-857 [FREE Full text] [doi: 10.1093/ije/dyq002] [Medline: 20207606]
13. Al Mutairi Sana S, Shihab-Eldeen AA, Mojiminiyi OA, Anwar AA. Comparative analysis of the effects of hubble-bubble (Sheesha) and cigarette smoking on respiratory and metabolic parameters in hubble-bubble and cigarette smokers. Respirology 2006 Jul;11(4):449-455. [doi: 10.1111/j.1440-1843.2006.00873.x] [Medline: 16771915]
14. Kiter G, Uçan ES, Ceylan E, Kilinç O. Water-pipe smoking and pulmonary functions. Respir Med 2000 Sep;94(9):891-894 [FREE Full text] [doi: 10.1053/rmed.2000.0859] [Medline: 11001082]
15. Inhorn MC, Buss KA. Ethnography, epidemiology and infertility in Egypt. Soc Sci Med 1994 Sep;39(5):671-686. [Medline: 7973866]
16. Monzer B, Sepetdjian E, Saliba N, Shihadeh A. Charcoal emissions as a source of CO and carcinogenic PAH in mainstream narghile waterpipe smoke. Food Chem Toxicol 2008 Sep;46(9):2991-2995. [doi: 10.1016/j.fct.2008.05.031] [Medline: 18573302]
17. WHO Study Group on Tobacco Product Regulation. Waterpipe Tobacco Smoking: Health Effects, Research Needs and Recommended Actions by Regulators. Geneva, Switzerland; 2005. URL: http://www.who.int/tobacco/global_interaction/tobreg/Waterpipe%20recommendation_Final.pdf [accessed 2015-06-09] [WebCite Cache ID 6ZA9RDzy4]
18. Fromme H, Dietrich S, Heitmann D, Dressel H, Diemer J, Schulz T, et al. Indoor air contamination during a waterpipe (narghile) smoking session. Food Chem Toxicol 2009 Jul;47(7):1636-1641. [doi: 10.1016/j.fct.2009.04.017] [Medline: 19394392]
19. Grekin ER, Ayna D. Waterpipe smoking among college students in the United States: A review of the literature. J Am Coll Health 2012;60(3):244-249. [doi: 10.1080/07448481.2011.589419] [Medline: 22420702]
20. Heinz AJ, Giedgowd GE, Crane NA, Veilleux JC, Conrad M, Braun AR, et al. A comprehensive examination of hookah smoking in college students: Use patterns and contexts, social norms and attitudes, harm perception, psychological correlates and co-occurring substance use. Addict Behav 2013 Nov;38(11):2751-2760. [doi: 10.1016/j.addbeh.2013.07.009] [Medline: 23934006]
21. Primack BA, Sidani J, Agarwal AA, Shadel WG, Donny EC, Eissenberg TE. Prevalence of and associations with waterpipe tobacco smoking among U.S. university students. Ann Behav Med 2008 Aug;36(1):81-86 [FREE Full text] [doi: 10.1007/s12160-008-9047-6] [Medline: 18719977]
22. Roskin J, Aveyard P. Canadian and English students' beliefs about waterpipe smoking: A qualitative study. BMC Public Health 2009;9:10 [FREE Full text] [doi: 10.1186/1471-2458-9-10] [Medline: 19134220]
23. Smith SY, Curbow B, Stillman FA. Harm perception of nicotine products in college freshmen. Nicotine Tob Res 2007 Sep;9(9):977-982. [doi: 10.1080/14622200701540796] [Medline: 17763115]
24. File T, Ryan C. Computer and Internet use in the United States: 2013. American Community Survey Reports (ACS-28). Washington, DC: US Census Bureau URL: http://www.census.gov/history/pdf/2013computeruse.pdf [accessed 2015-06-09] [WebCite Cache ID 6ZA9OIpAg]
25. Lenhart A, Purcell K, Smith A, Zickuhr K. Social Media and Mobile Internet Use Among Teens and Young Adults. Washington, DC: Pew Internet and American Life Project; 2010. URL: http://www.pewinternet.org/files/old-media/Files/Reports/2010/PIP_Social_Media_and_Young_Adults_Report_Final_with_toplines.pdf [accessed 2015-06-09] [WebCite Cache ID 6ZA9LQOLw]
26. Gross E. Adolescent Internet use: What we expect, what teens report. J Appl Dev Psychol 2004 Nov;25(6):633-649. [doi: 10.1016/j.appdev.2004.09.005]
27. Myslín M, Zhu S, Chapman W, Conway M. Using twitter to examine smoking behavior and perceptions of emerging tobacco products. J Med Internet Res 2013;15(8):e174 [FREE Full text] [doi: 10.2196/jmir.2534] [Medline: 23989137]

28. 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]
29. Whitsitt J, Karimkhani C, Boyers LN, Lott JP, Dellavalle RP. Comparing burden of dermatologic disease to search interest on google trends. Dermatol Online J 2015 Jan;21(1). [Medline: 25612125]
30. 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]
31. Nguyen T, Tran T, Luo W, Gupta S, Rana S, Phung D, et al. Web search activity data accurately predict population chronic disease risk in the USA. J Epidemiol Community Health 2015 Jul;69(7):693-699. [doi: 10.1136/jech-2014-204523] [Medline: 25805603]
32. Salloum RG, Osman A, Maziak W, Thrasher JF. How popular is waterpipe tobacco smoking? Findings from internet search queries. Tob Control 2015 Sep;24(5):509-513. [doi: 10.1136/tobaccocontrol-2014-051675] [Medline: 25052859]
33. Hicks A, Comp S, Horovitz J, Hovarter M, Miki M, Bevan J. Why people use Yelp.com: An exploration of uses and gratifications. Comput Hum Behav 2012 Nov;28(6):2274-2279. [doi: 10.1016/j.chb.2012.06.034]
34. Sussman S, Garcia R, Cruz TB, Baezconde-Garbanati L, Pentz MA, Unger JB. Consumers' perceptions of vape shops in Southern California: An analysis of online Yelp reviews. Tob Induc Dis 2014;12(1):22 [FREE Full text] [doi: 10.1186/s12971-014-0022-7] [Medline: 25484852]
35. 10 Things You Should Know About Yelp. URL: http://www.yelp.com/about [accessed 2015-06-09] [WebCite Cache ID 6ZA8HmAqJ]
36. Competitive Intelligence. Yelp.com Site Overview. URL: http://www.alexa.com/siteinfo/yelp.com [accessed 2015-06-09] [WebCite Cache ID 6ZA9Hpdel]
37. United States Census Bureau. 2014 U.S. Gazetteer files. [Data set]. 2014. URL: https://www.census.gov/geo/maps-data/data/gazetteer2014.html [accessed 2015-06-09] [WebCite Cache ID 6ZA9EvslA]
38. Buchholz N, Resnick S, Konty K. The New York City Community Health Survey Atlas, 2010. New York: The New York City Department of Health and Mental Hygeine; 2012. URL: http://www.nyc.gov/html/doh/downloads/pdf/epi/nyc_comhealth_atlas10.pdf [accessed 2015-03-30] [WebCite Cache ID 6ZA8BLQgr]
39. Noonan D. Exemptions for hookah bars in clean indoor air legislation: A public health concern. Public Health Nurs 2010;27(1):49-53. [doi: 10.1111/j.1525-1446.2009.00826.x] [Medline: 20055968]
40. Maziak W, Taleb ZB, Bahelah R, Islam F, Jaber R, Auf R, et al. The global epidemiology of waterpipe smoking. Tob Control 2015 Mar;24 Suppl 1:i3-i12 [FREE Full text] [doi: 10.1136/tobaccocontrol-2014-051903] [Medline: 25298368]
41. Barnett TE, Smith T, He Y, Soule EK, Curbow BA, Tomar SL, et al. Evidence of emerging hookah use among university students: A cross-sectional comparison between hookah and cigarette use. BMC Public Health 2013;13:302 [FREE Full text] [doi: 10.1186/1471-2458-13-302] [Medline: 23560649]
42. New York City Department of City Planning. Population Growth and Race/Hispanic Composition. New York: New York City Department of City Planning URL: http://www.nyc.gov/html/dcp/pdf/census/nyc20001.pdf [accessed 2015-06-09] [WebCite Cache ID 6ZA85aca7]
43. Museum N, Kayal P, York M, Benson K. A Community of Many Worlds: Arab Americans in New York City. New York: Museum of the City of New York; 2002.
44. Rice VH, Weglicki LS, Templin T, Hammad A, Jamil H, Kulwicki A. Predictors of Arab American adolescent tobacco use. Merrill Palmer Q (Wayne State Univ Press) 2006 Apr;52(2):327-342 [FREE Full text] [Medline: 16909165]
45. Lauterstein D, Hoshino R, Gordon T, Watkins B, Weitzman M, Zelikoff J. The changing face of tobacco use among United States youth. Curr Drug Abuse Rev 2014;7(1):29-43 [FREE Full text] [Medline: 25323124]
46. Primack BA, Aronson JD, Agarwal AA. An old custom, a new threat to tobacco control. Am J Public Health 2006 Aug;96(8):1339. [doi: 10.2105/AJPH.2006.090381] [Medline: 16809578]
47. Sterling KL, Mermelstein R. Examining hookah smoking among a cohort of adolescent ever smokers. Nicotine Tob Res 2011 Dec;13(12):1202-1209 [FREE Full text] [doi: 10.1093/ntr/ntr146] [Medline: 21896886]
48. Primack BA, Hopkins M, Hallett C, Carroll MV, Zeller M, Dachille K, et al. US health policy related to hookah tobacco smoking. Am J Public Health 2012 Sep;102(9):e47-e51 [FREE Full text] [doi: 10.2105/AJPH.2012.300838] [Medline: 22827447]
49. Kadry B, Chu LF, Kadry B, Gammas D, Macario A. Analysis of 4999 online physician ratings indicates that most patients give physicians a favorable rating. J Med Internet Res 2011;13(4):e95 [FREE Full text] [doi: 10.2196/jmir.1960] [Medline: 22088924]
50. Ellimoottil C, Kadlec AO, Farooq A, Quek ML. Choosing a physician in the Yelp era. Bull Am Coll Surg 2013 Nov;98(11):20-21. [Medline: 24313135]
51. Bakhsh W, Mesfin A. Online ratings of orthopedic surgeons: Analysis of 2185 reviews. Am J Orthop (Belle Mead NJ) 2014 Aug;43(8):359-363. [Medline: 25136868]
52. Booth D. Yelp partners with health departments to improve food safety. J Environ Health 2014 Apr;76(8):52-55. [Medline: 24749227]

53. 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]
54. The New York City Department of Health and Mental Hygiene. Undercover Health Department Investigation Found 13 Hookah Bars in Violation of the Smoke-Free Air Act. New York: The New York City Department of Health and Mental Hygiene; 2015 Jan 7. URL: http://www.nyc.gov/html/doh/html/pr2015/pr001-15.shtml [accessed 2015-06-09] [WebCite Cache ID 6ZA7xMpvV]

Edited by G Eysenbach; submitted 09.06.15; peer-reviewed by Y Wang, W Luo; comments to author 29.07.15; revised version received 09.08.15; accepted 22.09.15; published 20.11.15.

Please cite as:
Cawkwell PB, Lee L, Weitzman M, Sherman SE
Tracking Hookah Bars in New York: Utilizing Yelp as a Powerful Public Health Tool
JMIR Public Health Surveill 2015;1(2):e19
URL: http://publichealth.jmir.org/2015/2/e19/
doi:10.2196/publichealth.4809
PMID:27227137

Original Paper

A Comparison of Self-Reported and Objective Physical Activity Measures in Young Australian Women

Stefanie Hartley[1], BA, BSc, MPH; Suzanne Garland[1,2,3], MBBS, MD, FRCPA, FRANZCOG, FAChSHM; Elisa Young[1], BSc (Hons), MSc, PhD; Kim Louise Bennell[3], BAppSci, PhD; Ilona Tay[3], BSc, MSc; Alexandra Gorelik[3,4], BA, MSc; John Dennis Wark[3,4], MBBS, PhD, FRACP

[1]Murdoch Childrens Research Institute, Parkville, Australia

[2]The Royal Women's Hospital, Department of Molecular Microbiology & Infectious Diseases, Parkville, Australia

[3]University of Melbourne, Parkville, Australia

[4]Melbourne Health, Royal Melbourne Hospital, Parkville, Australia

Corresponding Author:
Suzanne Garland, MBBS, MD, FRCPA, FRANZCOG, FAChSHM
The Royal Women's Hospital
Department of Molecular Microbiology & Infectious Diseases
Locked Bag 3000
Parkville, 3052
Australia
Phone: 61 3 8345 3670
Fax: 61 3 9347 8235
Email: suzanne.garland@thewomens.org.au

Abstract

Background: The evidence for beneficial effects of recommended levels of physical activity is overwhelming. However, 70% of Australians fail to meet these levels. In particular, physical activity participation by women falls sharply between ages 16 to 25 years. Further information about physical activity measures in young women is needed. Self-administered questionnaires are often used to measure physical activity given their ease of application, but known limitations, including recall bias, compromise the accuracy of data. Alternatives such as objective measures are commonly used to overcome this problem, but are more costly and time consuming.

Objective: To compare the output between the Modified Active Australia Survey (MAAS), the International Physical Activity Questionnaire (IPAQ), and an objective physical activity measure—the SenseWear Armband (SWA)—to evaluate the test-retest reliability of the MAAS and to determine the acceptability of the SWA among young women.

Methods: Young women from Victoria, Australia, aged 18 to 25 years who had participated in previous studies via Facebook advertising were recruited. Participants completed the two physical activity questionnaires online, immediately before and after wearing the armband for 7 consecutive days. Data from the SWA was blocked into 10-minute activity times. Follow-up IPAQ, MAAS, and SWA data were analyzed by comparing the total continuous and categorical activity scores, while concurrent validity of IPAQ and MAAS were analyzed by comparing follow-up scores. Test-retest reliability of MAAS was analyzed by comparing MAAS total physical activity scores at baseline and follow-up. Participants provided feedback in the follow-up questionnaire about their experience of wearing the armband to determine acceptability of the SWA. Data analyses included graphical (ie, Bland-Altman plot, scatterplot) and analytical (ie, canonical correlation, kappa statistic) methods to determine agreement between MAAS, IPAQ, and SWA data.

Results: A total of 58 participants returned complete data. Comparisons between the MAAS and IPAQ questionnaires (n=52) showed moderate agreement for both categorical (kappa=.48, $P<.001$) and continuous data ($r=.69$, $P<.001$). Overall, the IPAQ tended to give higher scores. No significant correlation was observed between SWA and IPAQ or MAAS continuous data, for both minute-by-minute and blocked SWA data. The SWA tended to record lower scores than the questionnaires, suggesting participants tended to overreport their amount of physical activity. The test-retest analysis of MAAS showed moderate agreement for continuous outcomes ($r=.44$, $P=.001$). However, poor agreement was seen for categorical outcomes. The acceptability of the SWA to participants was high.

Conclusions: Moderate agreement between the MAAS and IPAQ and moderate reliability of the MAAS indicates that the MAAS may be a suitable alternative to the IPAQ to assess total physical activity in young women, due to its shorter length and consequently lower participant burden. The SWA, and likely other monitoring devices, have the advantage over questionnaires of avoiding overreporting of self-reported physical activity, while being highly acceptable to participants.

(JMIR Public Health Surveill 2015;1(2):e14) doi:10.2196/publichealth.4259

KEYWORDS

physical activity; exercise; women's health; questionnaires

Introduction

There is overwhelming evidence of the health, social, and economic benefits of engaging in recommended levels of physical activity [1,2]. Physical activity guidelines for Australians are to accumulate 150-300 minutes of moderate-intensity physical activity, 75-150 minutes of vigorous-intensity physical activity, or an equivalent combination of both moderate and vigorous activities each week [3]. However, close to three-quarters of the Australian population aged 18-65 years fail to meet these levels [4], a finding common to other developed countries [5]. Research into interventions to increase physical activity levels in the community is therefore warranted. In particular, young women should be targeted for interventions, as women play a large role in influencing household activity levels and the amount of physical activity women engage in falls sharply during the ages of 16-25 years [6].

Physical activity is defined as "any bodily movement produced by skeletal muscles that requires energy expenditure" [7]. It can occur in various forms and contexts, such as actions performed during recreation, sports, work, household chores, and gardening. The health benefits of physical activity include prevention and management of chronic diseases as well as overall reduced mortality and improved mental health [1,2].

Given the importance of physical activity, valid and precise assessment of activity levels are needed to determine activity trends, explore associations between health and physical activity, predict population health outcomes, and evaluate the effectiveness of physical activity interventions [8,9]. However, it is a methodological challenge to accurately measure physical activity in individuals, due to the complexities of daily life [10] and the technical requirements to make meaningful measurements. The burden and inadequate precision of quantifying physical activity levels has created a market for the development of better measuring methods. Various methods to capture levels of physical activity in free-living individuals are available and are categorized under (1) subjective (self-reported) and (2) objective measures. For research purposes, it is important to select reliable and valid measures that can be feasibly administered to individuals.

Self-reported questionnaires, while low cost and simple to administer, can vary in accuracy due to recall bias, social desirability bias, and misinterpretation [11]. A systematic review of physical activity questionnaires recommended 23 questionnaires, identified as having good content validity [8]. One of these was the International Physical Activity Questionnaire (IPAQ), which is considered to be the most extensively validated questionnaire across 12 countries [12]. The Modified Active Australia Survey (MAAS) is a less commonly used questionnaire to measure physical activity [13], although it has been employed in large Australian national and state surveys [13,14]. The MAAS was developed by shortening the Active Australia Survey, and has been shown to have comparable reliability and validity to the Active Australia Survey [13]. Both the IPAQ and MAAS ask the respondent about the duration, frequency, and intensity of activity in the preceding 7 days, and include activities such as walking, sports, yard and housework, and bicycling. MAAS has the advantage of being considerably shorter than the long-form IPAQ and is relevant to the Australian context. When tested in middle-aged Australian women, it has also been found to have comparable reliability and validity to that reported for the full version of the Active Australia Survey [13]. However, neither the test-retest reliability of the MAAS nor its concurrent validity compared to other questionnaires, such as the widely accepted IPAQ, has been established in younger women.

A variety of noninvasive objective measures is commercially available to assess physical activity levels. One of these is the SenseWear Armband (SWA) activity monitor (Model: MF-SW) (BodyMedia, Inc, Pittsburgh, PA, USA), which is a small, lightweight, multisensor activity monitor worn on the upper arm. It integrates body motion and step count from a three-axis accelerometer while other sensors such as heat flux, galvanic skin responses, and skin surface temperature can provide other data. The SWA has been extensively validated in numerous peer-reviewed publications [15]. However, the SWA is costly and not waterproof, meaning that it cannot record water-based physical activity. To date, a comparison of measures from the SWA and the IPAQ and MAAS has not been conducted in young Australian women. It is also not known whether the SWA is an acceptable method of collecting physical activity measures in this population.

The primary aims of the study in young Australian women were to (1) compare the output from the SWA with self-reported measures of physical activity by IPAQ and MAAS and (2) determine the acceptability of using the SWA to assess levels of physical activity. The secondary aims were to evaluate (1) the concurrent validity of MAAS by comparing IPAQ and MAAS and (2) the test-retest reliability of MAAS.

Methods

Participants

This project was a substudy of the Young Female Health Initiative (YFHI), a comprehensive study of lifestyle, health, and well-being in young women aged 16-25 years living in the state of Victoria, Australia. This study received ethical approval from the Human Research and Ethics Committees of the Royal Women's Hospital, Victoria, Australia.

Cross-recruitment of participants from previous studies, namely the YFHI pilot [16] and the Vaccine Against Cervical Cancer Impact aNd Effectiveness (VACCINE) studies [17], was the main recruitment method employed. The majority of women in these 2 studies had previously volunteered by responding to advertisements on Facebook. Women were approached for this study if they met the inclusion criteria and had consented to be contacted for future studies. Individuals who expressed an interest in the study from the YFHI website [18] and fulfilled the inclusion criteria also were recruited. The recruitment of 70 participants for this substudy took place from June to September 2012. All participants gave verbal and written informed consent for the study, after the nature and possible consequences were explained.

To be eligible for the substudy, participants needed to be YFHI study participants and to satisfy the following criteria: (1) female, (2) aged 16-25 years, (3) living in Victoria, Australia, (4) provide verbal and written consent, and (5) willing to complete 2 questionnaires and to wear an SWA for 7 consecutive days. Participants were excluded if they were living outside Victoria, were unable to give consent due to a language barrier, had a physical impairment, or if there were any other reasons that would affect the completion of the study.

Procedure

Participants were emailed a link to the online baseline questionnaire containing the MAAS (16 items) and IPAQ (self-administered long-form version; 49 items), and a set of questions on acceptability of the SWA. Questionnaires were administered using the online survey tool, SurveyMonkey [19].

Participants were sent a study package by post containing an information and consent form, an SWA with instructions on its use, and an armband usage log. Participants were instructed to wear the armband on the back of the upper left arm (over the triceps muscle) for 7 consecutive days, removing it only for water-based activities, such as showering or swimming. They were also asked to record on the monitor log provided when the SWA was removed, entering details including the time the armband was removed, the time it was replaced, activities undertaken during that time, and the intensity level of any physical activity (low, moderate, or vigorous). Participants were asked to return the package via registered post, or deliver it directly to the study office.

Participants were emailed a link to the follow-up physical activity questionnaires as completion of the 7-day period wearing the SWA approached. These were identical to the baseline questionnaires, with the addition of feedback questions about the participant's experience of wearing the SWA. Participants who completed the study were given AUD $10 in the form of a gift voucher to a retail store as minor compensation for their time.

Analysis of Questionnaires

MAAS asks the respondent the duration, frequency, and intensity of activity in leisure time, household or garden chores, and sedentary behavior in the last 7 days. The MAAS questionnaires were analyzed and scored according to the method described in Brown et al [13]. Participants who scored "none" in the total physical activity were grouped in the *low* category, to enable comparisons with IPAQ. The IPAQ scoring protocol was used for the IPAQ [20]. In brief, for both questionnaires, participants are classified into low, medium, and high levels of activity based on specified rules for occasions and minutes of various-intensity exercise. As well as this categorical outcome, a continuous outcome can be calculated based on metabolic equivalent of task (MET) minutes per week. Both MAAS and IPAQ give scores for different domains and intensities of physical activity. These are then combined to give a total score. For this study, only the total scores of the MAAS and IPAQ were used in the analysis.

Analysis of SenseWear Output

Returned devices were connected to a computer and minute-by-minute data were downloaded from the device using SenseWear Professional Software 7.0 (BodyMedia Inc, Pittsburgh, PA, USA). The data were used to estimate the frequency and duration of each activity as well as METs that are based on the participant's gender, age, height, and body mass. The MET data were analyzed twice as follows: (1) using minute-by-minute data, and (2) using *blocked* data. Blocked data were generated to allow comparisons between the SWA data and the 2 questionnaires. The IPAQ and MAAS ask only about physical activity performed for 10 minutes or more, whereas the SWA collects minute-by-minute data. To enable more meaningful comparisons between the SWA and the questionnaires, SWA data were blocked into activity of 10 minutes or more. A similar method of blocking was used by Brown et al [13].

The following conditions were applied to generate blocked data: (1) the time between the first and last reading must be 10 minutes or longer, (2) the first and last times must show an MET ≥ 3.3, (3) for blocks of 15 minutes or less, 80% of MET values must be ≥ 3.3, (4) for blocks of over 15 minutes, 75% of MET values must be ≥ 3.3, (5) for blocks of 15 minutes or less, "rests" must be no longer than 2 minutes, (6) for blocks of over 15 minutes, "rests" must be no longer than 3 minutes, (7) for blocks of 30-60 minutes, there can be one 4-minute "rest," (8) for blocks over 60 minutes, there can be two 4-minute "rests," and (9) rest was defined as an MET value of less than 3.3.

For our analyses, participants who had at least 4 monitoring days, including 2 weekend days, were included. A valid day was defined as having at least 1296 on-body minutes, after inclusion of known activities from the monitor log. This corresponds to 90% of a 24-hour period. Participants who fell short of the criteria were excluded from analysis.

Calculation of Scores

Average weekday scores were obtained by using the equation shown in Figure 1, where *x* denotes the number of valid monitoring weekdays the participant has provided. A similar method was used to generate average weekend scores.

With the averages for both weekday and weekend scores, a weekly total score was then calculated using the following formula:

5 (average weekday score) + 2 (average weekend score)

Figure 1. Equation used to calculate average weekday scores from the SenseWear Armband activity monitor.

$$\sum \frac{walking\ minutes}{\frac{x\ weekdays}{3.3}} + \sum \frac{moderate\ minutes}{\frac{x\ weekdays}{4.0}} + \sum \frac{vigorous\ minutes}{\frac{x\ weekdays}{8.0}}$$

Statistical Analysis

Data analyses were performed using SPSS version 17.0 (SPSS Inc, Chicago, IL, USA) and STATA version 13.0 (StataCorp LP, College Station, TX, USA). Descriptive statistics were reported as n and percentages for categorical data, and mean (SD) for continuous data, except the interval between completion of baseline and follow-up questionnaires that was reported as median (interquartile range [IQR]). All continuous data were tested for normality prior to data analysis. Both MAAS and IPAQ scores had skewed distributions and were transformed using natural logarithmic transformation. MAAS and IPAQ were analyzed twice—as continuous and as categorical.

Pearson correlation was used to measure the association between total continuous outcomes of the self-reported questionnaires (IPAQ and MAAS) with the SWA. Calculation of MAAS test-retest reliability was only conducted for participants who indicated that their overall physical activity levels had not changed between completion of the baseline and follow-up questionnaire, due to wearing the SWA.

Standard analytical (ie, canonical correlation) and graphical (ie, scatterplot, Bland-Altman plot) techniques were used to determine the agreement between continuous measures—MAAS, IPAQ, and SWA scores—while kappa statistics were used to determine the agreement between categorical data—IPAQ and MAAS categories. In all cases, statistical significance was defined at $P<.05$.

Results

Participants

A total of 58 participants returned an SWA with data. Of these, 54 (93%) completed the baseline questionnaire and 52 (90%) completed the follow-up questionnaire. A total of 4 out of 58 (7%) were excluded as they had less than 4 days of recorded data. The mean days of SWA wear was 6.43 (SD 0.67; range 4-7) and the mean total time of SWA wear was 133.65 hours (SD 21.01; range 68.15-161.57). The mean age of participants was 22.1 years (SD 2.0; range 18.5-25.3). Demographic characteristics of participants are shown in Table 1.

Table 1. Demographic characteristics of participants (n=54).

Characteristic	Categories	n (%)[a]
Age (years)		
	18-21	26 (48)
	22-25	28 (52)
Country of birth		
	Australia	52 (96)
	Other	2 (4)
Geographic region of residence		
	Major city	45 (83)
	Inner regional	9 (17)
	Outer regional/remote	0 (0)
Body mass index (kg/m ^{2}) (using self-reported height and weight)		
	Underweight	6 (11)
	Normal	32 (59)
	Overweight	13 (24)
	Obese	3 (6)
	Extremely obese	0 (0)
Education level [b]		
	<Year 12	4 (7)
	Year 12	19 (35)
	>Year 12	31 (57)
Socioeconomic level (SEIFA [c] **percentile)** [d]		
	≤25 (most disadvantaged)	2 (4)
	26-100	52 (96)

[a]Percentages may not add to exactly 100 due to rounding.

[b]Year 12 is the final year of high school in the Australian education system.

[c]Socio-Economic Indexes For Areas (SEIFA).

[d]Based on postal/zip code. Percentiles are the rankings within Victoria, with a percentile of ≤25 being the most disadvantaged quartile.

Criterion Validity of the International Physical Activity Questionnaire and Modified Active Australia Survey

There was no significant correlation observed between follow-up IPAQ scores and SWA minute-by-minute continuous data (r=.10, P=.48) and follow-up IPAQ scores and SWA blocked continuous data (r=.07, P=.63). Both distributions are scattered and no linearity is observed. The negative slope of the band in the Bland-Altman plot reveals a tendency for IPAQ to give higher scores than the SWA (see Figure 2).

No significant correlations were observed between follow-up MAAS scores and SWA minute-by-minute continuous data and SWA blocked continuous data (r=.02, P=.88; r=.05, P=.72, respectively; Figure 3).

Figure 2. Bland-Altman plot of natural log transformed score showing the difference between IPAQ and SWA, plotted against the mean. Note: the lines represent the limits of agreement (95%) and average difference between the two variables.

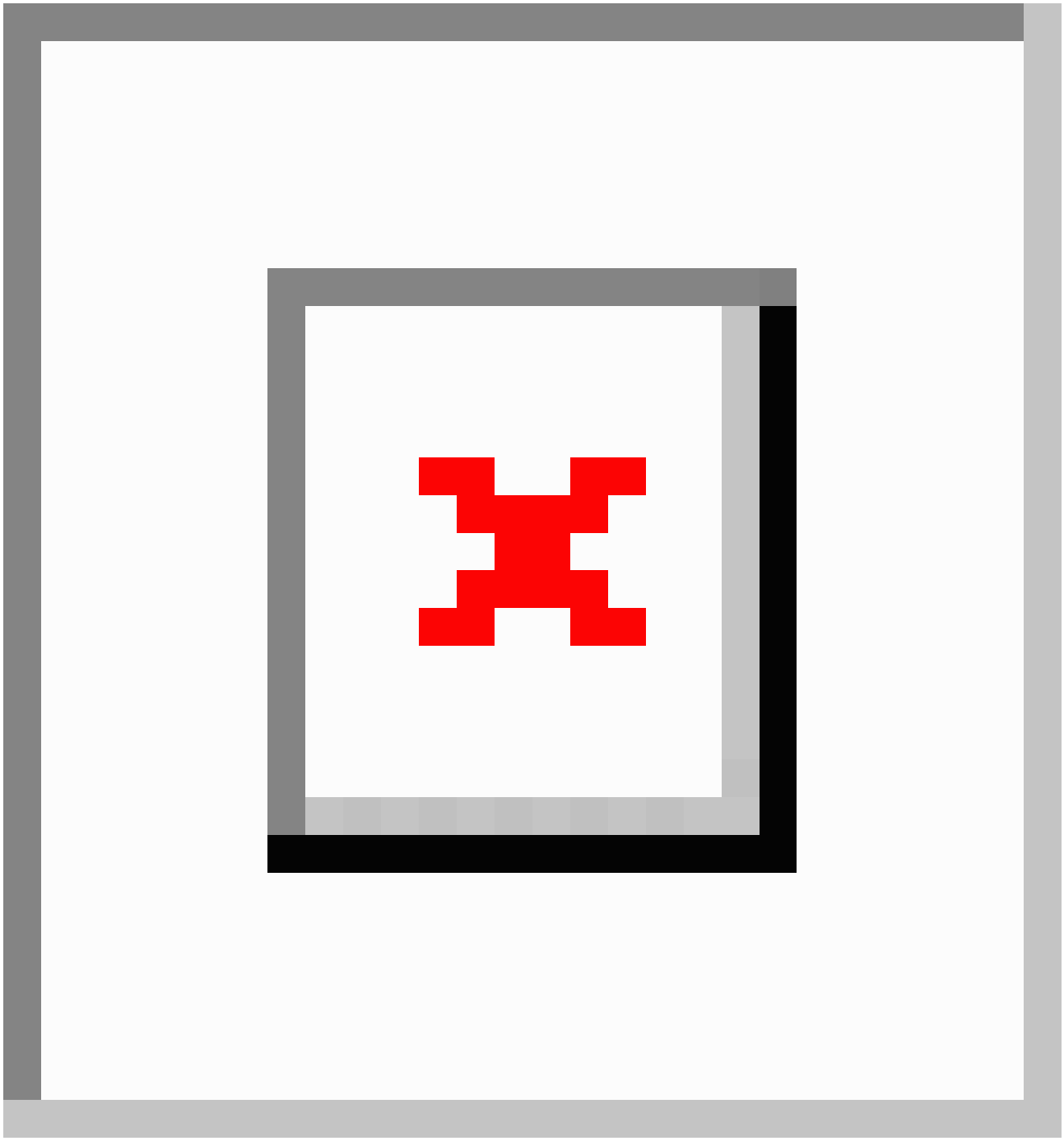
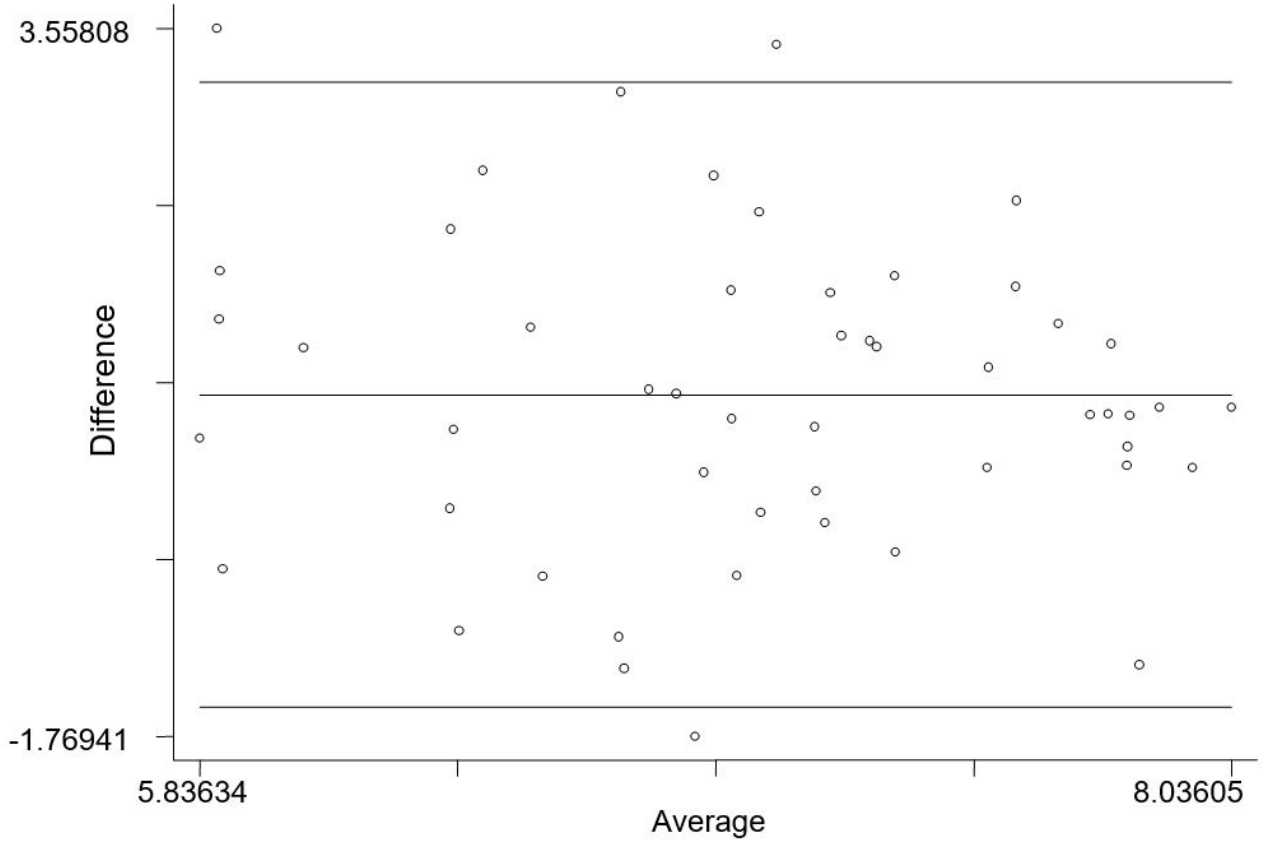


Figure 3. Bland-Altman plot of natural log transformed score showing the difference between MAAS and SWA, plotted against the mean. Note: the lines represent the limits of agreement (95%) and average difference between the two variables.



Acceptability of the SenseWear Armband

Table 2 summarizes participants' experiences using the SWA. The majority of participants reported a positive experience, with more than half (29/52, 56%) reporting the SWA as not painful to wear, and just under a third (15/52, 29%) reporting not feeling at all self-conscious wearing it. A large proportion of participants (>75%) did not stop wearing the SWA due to feeling self-conscious about wearing it, finding it uncomfortable or painful to wear, interfering with sleep, being prohibited from wearing it, or concerns of getting it wet. A total of 67% (35/52) of participants preferred using the SWA over completing a survey about their physical activity, and 29% (15/52) were interested in using the SWA daily on an ongoing basis. The majority of participants (34/52, 65%) reported no increase in physical activity because of wearing the SWA.

Table 2. Participants' experiences of using the SenseWear Armband (n=52)[a].

Statements from the questionnaire	Agreement scale (1=completely false to 5=completely true), n (%)					Median (IQR[b])
	1	2	3	4	5	
I often felt self-conscious wearing the activity monitor.	15 (29)	11 (21)	13 (25)	11 (21)	2 (4)	2.5 (1.0-3.3)
I often found the activity monitor uncomfortable to wear.	10 (19)	17 (33)	14 (27)	7 (13)	4 (8)	2 (2-3)
I often found the activity monitor painful to wear.	29 (56)	13 (25)	7 (13)	2 (4)	1 (2)	1 (1-2)
I often felt proud to be seen wearing the activity monitor.	7 (13)	17 (33)	18 (35)	9 (17)	1 (2)	3 (2-3)
I often did not wear the activity monitor because I felt self-conscious about being seen wearing it.	43 (83)	7 (13)	1 (2)	1 (2)	0 (0)	1 (1-1)
I often did not wear the activity monitor because it was uncomfortable or painful.	42 (81)	7 (13)	3 (6)	0 (0)	0 (0)	1 (1-1)
I often did not wear the activity monitor overnight because it interfered with my sleep.	40 (77)	5 (10)	4 (8)	2 (4)	1 (2)	1 (1-1)
I often did not wear the activity monitor during exercise because it would have got wet (eg, while swimming, or because I was walking in the rain).	37 (71)	7 (13)	2 (4)	3 (6)	3 (6)	1 (1-2)
I often did not wear the activity monitor because it was prohibited (eg, not allowed to wear it during netball, or in my workplace).	46 (88)	3 (6)	0 (0)	2 (4)	1 (2)	1 (1-1)
I exercised more than I otherwise would have, because of wearing the activity monitor.	34 (65)	13 (25)	3 (6)	1 (2)	1 (2)	1 (1-2)
I would prefer to record my physical activity for the last 7 days by completing a survey rather than wearing the activity monitor for 7 days.	35 (67)	7 (13)	7 (13)	2 (4)	1 (2)	1 (1-2)
I would like to wear the activity monitor everyday if I could get real-time feedback of my physical activity and calories burned.	5 (10)	4 (8)	13 (25)	15 (29)	15 (29)	4 (3-5)

[a]There were 52 respondents due to missing data on these items for 2 participants.

[b]Interquartile range (IQR).

Concurrent Validity of the Modified Active Australia Survey

Both categorical and continuous measures of the follow-up MAAS and IPAQ showed moderate agreement. Based on categorical data, 65% of cases (34/52) were in agreement, which corresponds to moderate agreement between 2 measures (kappa=.48, P<.001) (see Table 3).

Table 3. Categorical outcomes from MAAS[a]and IPAQ[b]follow-up questionnaires.

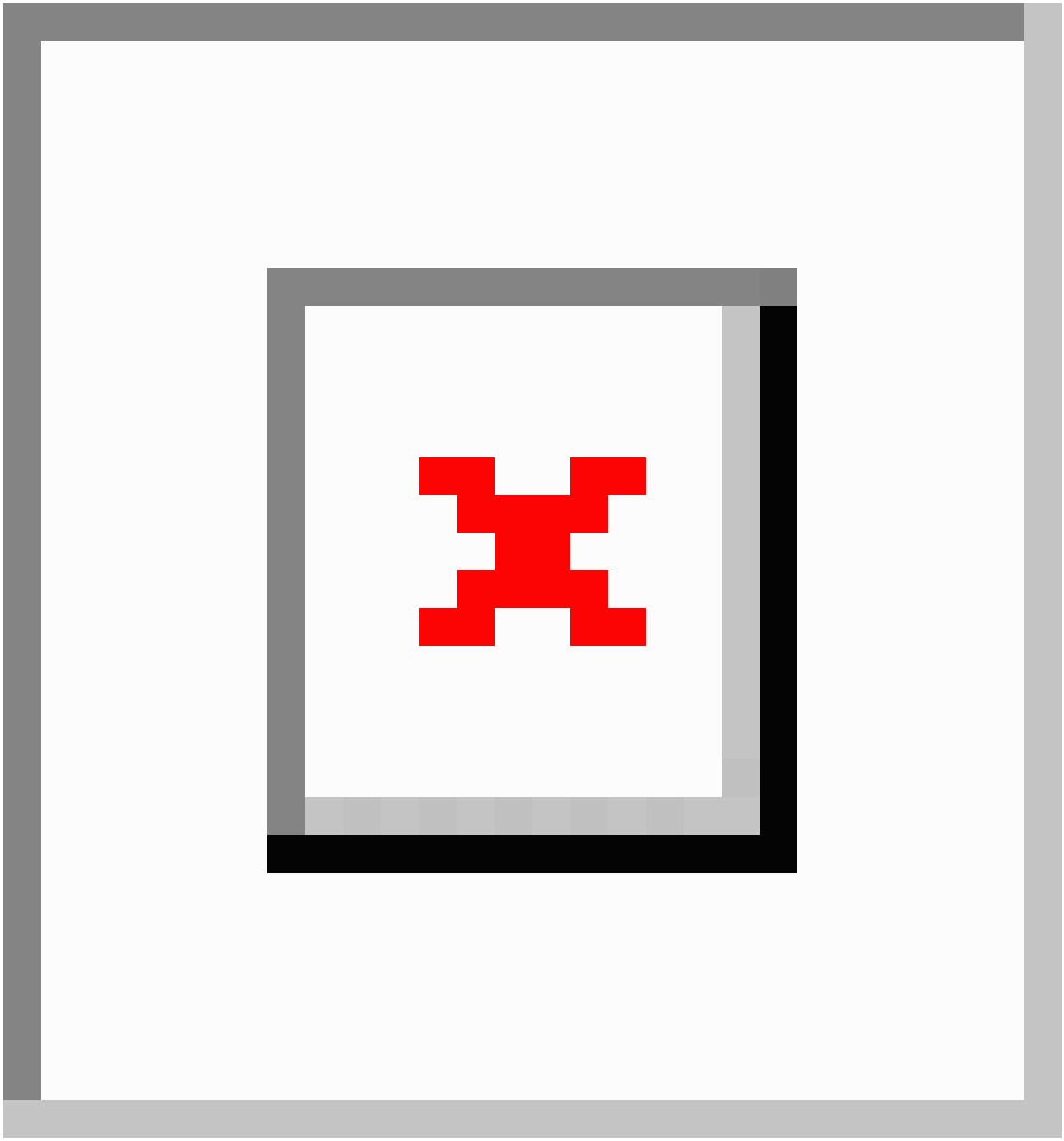
		IPAQ, n (%)			
		Low (n=17)	Moderate (n=16)	High (n=19)	Total (n=52)
MAAS, n (%)					
	Low	13 (76)	3 (19)	1 (5)	17 (33)
	Moderate	4 (24)	8 (50)	5 (26)	17 (33)
	High	0 (0)	5 (31)	13 (68)	18 (35)
	Total	17 (100)	16 (100)	19 (100)	52 (100)

[a]Modified Active Australia Survey (MAAS).

[b]International Physical Activity Questionnaire (IPAQ).

The results are similar when analyzed using continuous data with a canonical correlation of .69 (P<.001) and are in agreement with no visible consistent bias between 2 scores (limits of agreement are between -0.78 and 2.45; Figure 4).

Figure 4. Bland-Altman plot of natural log transformed score showing the difference between follow-up IPAQ and MAAS, plotted against the mean. Note: the lines represent the limits of agreement (95%) and average difference between the two variables.



Test-Retest Analysis of the Modified Active Australia Survey

The median interval between completion of baseline and follow-up questionnaires was 11.5 days (IQR 9.0-14.5). In analysis of test-retest reliability of the MAAS, 2 participants were excluded as they reported increasing their activity levels as a result of using the SWA. There was poor agreement between baseline and follow-up MAAS categorical scores (kappa=.193, *P*=.03; Table 4).

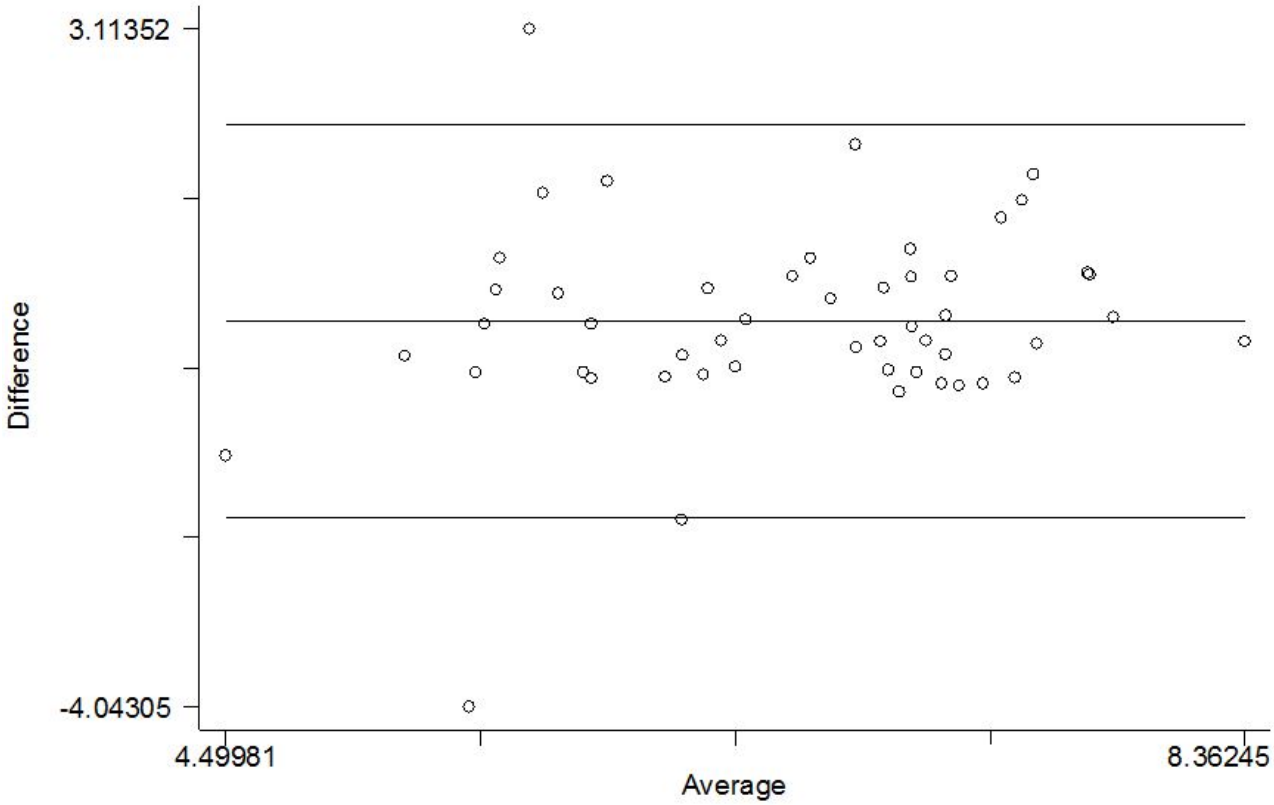
Table 4. Categorical outcomes from baseline and follow-up MAAS[a].

		Follow-up MAAS category, n (%)			
		Low (n=17)	Moderate (n=17)	High (n=18)	Total (n=52)
Baseline MAAS category, n (%)					
	Low	11 (65)	5 (29)	2 (11)	18 (35)
	Moderate	4 (24)	5 (29)	8 (44)	17 (33)
	High	2 (12)	7 (41)	8 (44)	17 (33)
	Total	17 (100)	17 (100)	18 (100)	52 (100)

[a]Modified Active Australia Survey (MAAS).

Canonical correlation yielded a correlation coefficient of .44 (*P*=.001). To show the differences in score as a function of the mean score, a Bland-Altman plot was constructed of the naturally logged baseline and follow-up MAAS scores (Figure 5). There was a mean difference of 0.03 (95% limits of agreement -2.04 to 2.10).

Figure 5. Bland-Altman plot of natural log transformed score showing the difference between baseline and follow-up MAAS, plotted against the mean. Note: the lines represent the limits of agreement (95%) and average difference between the two variables.



Discussion

Principal Findings

This study examined clinimetric properties of 3 physical activity assessment tools in young Australian women. The results showed that there was no significant correlation between SWA and IPAQ continuous data, either minute-by-minute or blocked SWA data. Comparison of the SWA and MAAS showed the same finding. The SWA tended to record lower scores than IPAQ and MAAS, suggesting participants tended to overreport their amount of physical activity. This effect is likely larger than shown in the results, as the SWA also captures incidental exercise, whereas IPAQ and MAAS do not. The amount of overreport is therefore likely larger and more substantial than observed. This finding of participant overreporting when compared to an objective monitoring device is consistent with findings from the much larger National Health And Nutrition Examination Survey (NHANES), which collected data from males and females aged 20 to 80+ years [21]. Physical activity monitors also have limitations in capturing all activity, and so may give an underreport of total activity which contributes to the differences between the SWA and questionnaire data.

Overall, the acceptability of the SWA was high. Notably, the majority of participants preferred to use the SWA rather than self-report their physical activity. A small number of participants (3/52, 6%) experienced localized skin reactions to the monitor,

a known rare complication due to metal allergy. Despite this, adherence to SWA use was high and participant responses to the feedback questions indicated that the SWA is both acceptable and favored by young Australian women. While adherence to the SWA was high, it is not known whether repeated use would give similar adherence levels. It may be that the young women found the SWA to be a novelty initially and so were pleased to wear it for 7 days, but that enthusiasm may wane with repeated use. Given that studies often require participants to collect physical activity data more than once over an extended period, it would be important to ascertain young women's preference for the SWA or questionnaires with repeated use.

According to Scheers et al [11], participants ideally should have at least 5 monitoring days, including 2 weekend days. However, for our analyses participants who had at least 4 monitoring days, including 2 weekend days, were included. This decision was made due to participant numbers. A larger cohort size would have allowed the conditions of Scheers et al to be applied, without losing too many participants from analysis. This is a limitation of this study which needs to be considered in interpreting the findings. It is possible that the days the participants wore the SWA were their most or least active days of the week, which would give an overestimation or underestimation of their total weekly activity, respectively. However, the agreement of our findings with the NHANES study [21], as discussed above, gives us greater confidence in them, despite this limitation.

Since this study was conducted, there has been a large increase in the availability and range of devices available for objectively monitoring physical activity and other body parameters, such as sleep and calorie intake. Many of these have the advantage of being considerably less costly and more lightweight than the SWA, and future studies would need to consider the advantages of using the SWA compared with these newer devices. However, our finding that a monitoring device tends to record lower levels of physical activity than self-reported questionnaires is likely applicable to other devices. We would expect that these devices would also share the acceptability to young women found for the SWA in this study, and that they may indeed be more acceptable due to their smaller and more lightweight design.

The test-retest reliability of the MAAS was moderate when using continuous outcomes. This is consistent with data from a study using MAAS in 159 middle-aged Australian women [13]. However, test-retest reliability was poor when using categorical outcomes. A possible explanation for lack of strong agreement is that participants were not excluded from this analysis if they indicated their activity levels had changed due to factors other than wearing the SWA. In order to be excluded from this analysis, participants had to indicate that their level of activity had increased due to wearing the SWA. Participants were not asked about increased or decreased physical activity due to factors other than the SWA—such as illness, holidays, or work commitments—so participants who experienced this were not excluded from this analysis.

Comparisons between the MAAS and IPAQ questionnaires showed moderate agreement for both categorical and continuous data. There was an overall tendency for the IPAQ to give a higher score, which may be due to the more detailed questions on different domains of activity contained in the IPAQ. Given the moderate level of agreement between MAAS and IPAQ and the moderate test-retest reliability when using continuous outcomes, MAAS may be a suitable alternative to IPAQ to assess total physical activity score, due to its shorter length and consequent reduction in participant burden. The analyses performed in this study were only for the total physical activity scores obtained using each instrument. Further analyses of the correlation between domains of physical activity would be useful in assessing comparison of the instruments.

A limitation of this study is the relatively small sample size and this needs to be considered in interpreting the results. A larger study would allow our findings to be further tested. The recruitment method used in this study has been shown to recruit a representative sample of young women, therefore we are confident our sample was not overly selective [16].

Conclusions

Physical activity is an important contributor to health and the prevention of disease, and its study requires reliable and valid measures which can be administered to individuals. This study showed that the MAAS has moderate agreement with the IPAQ, and as such may be a suitable alternative in some situations. Test-retest analysis of the MAAS did not give strong results; however, this is potentially explained by a limitation of the study in the omission of a question on changes in activity levels. This study showed that young women tended to overreport their physical activity when compared to the SWA monitor, suggesting the SWA may be a more accurate tool which overcomes some of the limitations of self-report instruments. This finding, combined with the high acceptability of the device to young women, suggests the SWA, and perhaps other such monitors, may be a better option for measuring physical activity for both researchers and research participants.

Acknowledgments

The researchers acknowledge Avelyn Yuan Hui Sim, whose Bachelor of Biomedicine (Honors) thesis contributed to the work in this manuscript. We would like to thank all the young women who participated in this study and joined us in our commitment to improving women's health. We would also like to acknowledge the wider YFHI study group for their support and contributions.

Conflicts of Interest

Professor Suzanne Garland has received grant support to her institution from CSL Bio, Merck, and GlaxoSmithKline (GSK) as well as receiving funding through her institution to conduct human papillomavirus (HPV) vaccine studies for the Merck Sharp

& Dohme (MSD) GSK Human PapillomaVIrus Vaccine Immunogenicity ANd Efficacy (VIVIANE) trial. Professor Garland is a member of the Merck Global Advisory Board as well as the Merck Scientific Advisory Committee for HPV (unpaid position).

References

1. Warburton DE, Nicol CW, Bredin SS. Health benefits of physical activity: the evidence. CMAJ 2006 Mar 14;174(6):801-809 [FREE Full text] [doi: 10.1503/cmaj.051351] [Medline: 16534088]
2. Fox KR. The influence of physical activity on mental well-being. Public Health Nutr 1999 Sep;2(3A):411-418. [Medline: 10610081]
3. Brown WJ, Bauman AE, Bull FC, Burton NW. Development of Evidence-Based Physical Activity Recommendations for Adults (18-64 Years). Canberra, Australia: Australian Government Department of Health; 2012 Aug 01. URL: http://www.health.gov.au/internet/main/publishing.nsf/Content/health-pubhlth-strateg-phys-act-guidelines/$File/DEB-PAR-Adults-18-64years.pdf [accessed 2015-09-29] [WebCite Cache ID 6buH1j5kK]
4. Brown H, Roberts J. Exercising choice: the economic determinants of physical activity behaviour of an employed population. Soc Sci Med 2011 Aug;73(3):383-390 [FREE Full text] [doi: 10.1016/j.socscimed.2011.06.001] [Medline: 21757272]
5. Hallal PC, Andersen LB, Bull FC, Guthold R, Haskell W, Ekelund U, Lancet Physical Activity Series Working Group. Global physical activity levels: surveillance progress, pitfalls, and prospects. Lancet 2012 Jul 21;380(9838):247-257. [doi: 10.1016/S0140-6736(12)60646-1] [Medline: 22818937]
6. Christie J, Nowson C, Garland S, Wark J. Emerging nutritional and lifestyle risk factors for bone health in young women: a mixed longitudinal twin study. In: Burckhardt P, Dawson-Hughes B, Weaver CM, editors. Nutritional Influences on Bone Health: 8th International Symposium. London, UK: Springer-Verlag; 2013:61-69.
7. Caspersen CJ, Powell KE, Christenson GM. Physical activity, exercise, and physical fitness: definitions and distinctions for health-related research. Public Health Rep 1985 Apr;100(2):126-131 [FREE Full text] [Medline: 3920711]
8. van Poppel MN, Chinapaw MJ, Mokkink LB, van Mechelen W, Terwee CB. Physical activity questionnaires for adults: a systematic review of measurement properties. Sports Med 2010 Jul 1;40(7):565-600. [doi: 10.2165/11531930-000000000-00000] [Medline: 20545381]
9. Chinapaw MJ, Mokkink LB, van Poppel MN, van Mechelen W, Terwee CB. Physical activity questionnaires for youth: a systematic review of measurement properties. Sports Med 2010 Jul 1;40(7):539-563. [doi: 10.2165/11530770-000000000-00000] [Medline: 20545380]
10. Zhang K, Pi-Sunyer FX, Boozer CN. Improving energy expenditure estimation for physical activity. Med Sci Sports Exerc 2004 May;36(5):883-889. [Medline: 15126725]
11. Scheers T, Philippaerts R, Lefevre J. Variability in physical activity patterns as measured by the SenseWear Armband: how many days are needed? Eur J Appl Physiol 2012 May;112(5):1653-1662. [doi: 10.1007/s00421-011-2131-9] [Medline: 21874552]
12. Craig CL, Marshall AL, Sjöström M, Bauman AE, Booth ML, Ainsworth BE, et al. International physical activity questionnaire: 12-country reliability and validity. Med Sci Sports Exerc 2003 Aug;35(8):1381-1395. [doi: 10.1249/01.MSS.0000078924.61453.FB] [Medline: 12900694]
13. Brown WJ, Burton NW, Marshall AL, Miller YD. Reliability and validity of a modified self-administered version of the Active Australia physical activity survey in a sample of mid-age women. Aust N Z J Public Health 2008 Dec;32(6):535-541. [doi: 10.1111/j.1753-6405.2008.00305.x] [Medline: 19076744]
14. Armstrong T, Bauman A, Davies J. Physical Activity Patterns of Australian Adults: Results of the 1999 National Physical Activity Survey. Canberra, Australia: Australian Institute of Health and Welfare; 2000. URL: http://www.aihw.gov.au/WorkArea/DownloadAsset.aspx?id=6442454841 [accessed 2015-09-29] [WebCite Cache ID 6buHpw4qn]
15. Liden CB, Wolowicz M, Stivoric J, Teller A, Vishnubhatla S, Pelletier R, et al. Integrated Fitness Systems. Benefits of the SenseWear Armband over other physical activity and energy expenditure measurement techniques URL: http://www.integratedfitnesssystems.com/wp-content/uploads/2012/10/Benefits.pdf [accessed 2015-01-20] [WebCite Cache ID 6VihQiQZo]
16. Fenner Y, Garland SM, Moore EE, Jayasinghe Y, Fletcher A, Tabrizi SN, et al. Web-based recruiting for health research using a social networking site: an exploratory study. J Med Internet Res 2012;14(1):e20 [FREE Full text] [doi: 10.2196/jmir.1978] [Medline: 22297093]
17. Young EJ, Tabrizi SN, Brotherton JM, Wark JD, Pyman J, Saville M, et al. Measuring effectiveness of the cervical cancer vaccine in an Australian setting (the VACCINE study). BMC Cancer 2013;13:296 [FREE Full text] [doi: 10.1186/1471-2407-13-296] [Medline: 23777549]
18. Young Female Health Initiative. URL: https://www.yfhi.org/ [accessed 2015-09-16] [WebCite Cache ID 6baf35N3p]
19. SurveyMonkey. URL: https://www.surveymonkey.com/ [accessed 2015-09-16] [WebCite Cache ID 6bafnhYwj]
20. Guidelines for Data Processing and Analysis of the International Physical Activity Questionnaire (IPAQ): Short and Long Forms.: International Physical Activity Questionnaire; 2005 Nov. URL: http://www.ipaq.ki.se/scoring.pdf [accessed 2015-01-20] [WebCite Cache ID 6Vihh9xAX]

21. Schuna JM, Johnson WD, Tudor-Locke C. Adult self-reported and objectively monitored physical activity and sedentary behavior: NHANES 2005-2006. Int J Behav Nutr Phys Act 2013;10:126 [FREE Full text] [doi: 10.1186/1479-5868-10-126] [Medline: 24215625]

Abbreviations

GSK: GlaxoSmithKline
HPV: human papillomavirus
IPAQ: International Physical Activity Questionnaire
IQR: interquartile range
MAAS: Modified Active Australia Survey
MET: metabolic equivalent of task
MSD: Merck Sharp & Dohme
NHANES: National Health And Nutrition Examination Survey
SEIFA: Socio-Economic Indexes For Areas
SWA: SenseWear Armband
VACCINE: Vaccine Against Cervical Cancer Impact aNd Effectiveness
VIVIANE: Human PapillomaVIrus Vaccine Immunogenicity ANd Efficacy
YFHI: Young Female Health Initiative

Edited by G Eysenbach; submitted 19.01.15; peer-reviewed by A Reimers, M van Poppel, R Joseph, D Pekmezi; comments to author 25.05.15; revised version received 13.07.15; accepted 15.08.15; published 05.10.15.

Please cite as:
Hartley S, Garland S, Young E, Bennell KL, Tay I, Gorelik A, Wark JD
A Comparison of Self-Reported and Objective Physical Activity Measures in Young Australian Women
JMIR Public Health Surveill 2015;1(2):e14
URL: http://publichealth.jmir.org/2015/2/e14/
doi:10.2196/publichealth.4259
PMID:27227132

Original Paper

Machine Translation of Public Health Materials From English to Chinese: A Feasibility Study

Anne M Turner[1*], MD, MLIS, MPH; Kristin N Dew[2*], MS; Loma Desai[3], MS, MBA; Nathalie Martin[4*], BA; Katrin Kirchhoff[5*], PhD

[1]Northwest Center for Public Health Practice, Department of Health Services, University of Washington, Seattle, WA, United States

[2]Northwest Center for Public Health Practice, Human Centered Design & Engineering, University of Washington, Seattle, WA, United States

[3]Northwest Center for Public Health Practice, Information School, University of Washington, Seattle, WA, United States

[4]Northwest Center for Public Health Practice, Department of Biomedical Informatics and Medical Education, University of Washington, Seattle, WA, United States

[5]Speech, Signal and Language Interpretation (SSLI) Lab, Department of Electrical Engineering, University of Washington, Seattle, WA, United States

[*]these authors contributed equally

Corresponding Author:
Anne M Turner, MD, MLIS, MPH
Northwest Center for Public Health Practice
Department of Health Services
University of Washington
Suite 400
1107 NE 45th Street
Seattle, WA, 98105
United States
Phone: 1 206 491 1489
Fax: 1 206 616 5249
Email: amturner@uw.edu

Abstract

Background: Chinese is the second most common language spoken by limited English proficiency individuals in the United States, yet there are few public health materials available in Chinese. Previous studies have indicated that use of machine translation plus postediting by bilingual translators generated quality translations in a lower time and at a lower cost than human translations.

Objective: The purpose of this study was to investigate the feasibility of using machine translation (MT) tools (eg, Google Translate) followed by human postediting (PE) to produce quality Chinese translations of public health materials.

Methods: From state and national public health websites, we collected 60 health promotion documents that had been translated from English to Chinese through human translation. The English version of the documents were then translated to Chinese using Google Translate. The MTs were analyzed for translation errors. A subset of the MT documents was postedited by native Chinese speakers with health backgrounds. Postediting time was measured. Postedited versions were then blindly compared against human translations by bilingual native Chinese quality raters.

Results: The most common machine translation errors were errors of word sense (40%) and word order (22%). Posteditors corrected the MTs at a rate of approximately 41 characters per minute. Raters, blinded to the source of translation, consistently selected the human translation over the MT+PE. Initial investigation to determine the reasons for the lower quality of MT+PE indicate that poor MT quality, lack of posteditor expertise, and insufficient posteditor instructions can be barriers to producing quality Chinese translations.

Conclusions: Our results revealed problems with using MT tools plus human postediting for translating public health materials from English to Chinese. Additional work is needed to improve MT and to carefully design postediting processes before the MT+PE approach can be used routinely in public health practice for a variety of language pairs.

(JMIR Public Health Surveill 2015;1(2):e17) doi:10.2196/publichealth.4779

KEYWORDS

public health informatics; public health; natural language processing; machine translation; Chinese language; health promotion; public health departments; consumer health; limited English proficiency; health literacy

Introduction

A key role of public health departments is to inform and educate the public on issues of public health importance. Health departments produce health promotion materials on a range of topics, such as environmental health, communicable diseases, immunizations, and maternal-child health, and the Internet has become a key mechanism by which they distribute and disseminate this information. Although federal and state regulations require that health materials be made available in the languages of patients, due to the time and costs required to manually produce quality translations, very few of these materials are available in languages other than English [1]. Therefore, individuals with limited English proficiency (LEP) have limited access to this health information. This is of particular significance given that LEP status is associated with poor health literacy and negative health consequences, including documented health disparities such as poorer health outcomes and poorer access to health care and preventive services compared to English-speaking minorities [2-4].

Machine translation (MT)—the automatic translation of text from one human language into another by a computer program—has been an area of study within natural language processing for several decades. State-of-the-art MT tools use a statistical machine translation (SMT) framework. This approach uses large amounts of parallel text for the desired language pair to train SMT models. During testing, an SMT engine then produces the most likely translation under the statistical model. While MT tools have improved greatly over the last 5 years, and MT is now routinely used by many language service providers, the quality of raw MT output generally falls short of human-generated translations (HT).

In order to produce quality translations, MT errors need to be corrected by human readers who have domain expertise and are fluent in the source and target languages. This correction, called postediting (PE), can range from light to heavy editing. It has been shown that MT+PE increases productivity (ie, it can be completed more quickly than producing an entirely new HT) both for translators and for lay users [5]. However, compared with translating, postediting is a cognitively different process, and postediting results are strongly dependent on posteditor skill, attitudes towards machine translation, difficulty of the source document, and quality of the initial machine translation output [5,6].

Our previous research indicates that freely available MT tools, such as Google Translate and Microsoft Translator, can be used in conjunction with human PE to produce quality translations efficiently and at low cost [7,8]. We compared the time and cost of HT versus MT+PE for Spanish public health documents, using health professionals as posteditors [7]. Posteditors corrected 25 machine-translated public health documents. Pairs of HT and MT+PE were blindly presented to 2 bilingual public health professionals, who were asked to rate which of the translations they preferred. In this blinded rating, the HT and MT+PE were found to be overall equivalent (33% HT preferred, 33% MT+PE preferred, 33% both translations considered equivalent).

These previous studies were conducted on a single language pair of English-Spanish. SMT generally works best when the source and target languages have similar sentence structures, as in the case of English-Spanish. In order to assess the broader usefulness of MT technology in public health departments, it is necessary to determine whether these results generalize to a wider set of language pairs, specifically those pairs with very divergent linguistic structures. One such pair, English-Chinese, is of particular interest since Chinese is the second most common language spoken by LEP individuals in the United States, representing 6.1% of the LEP population [9].

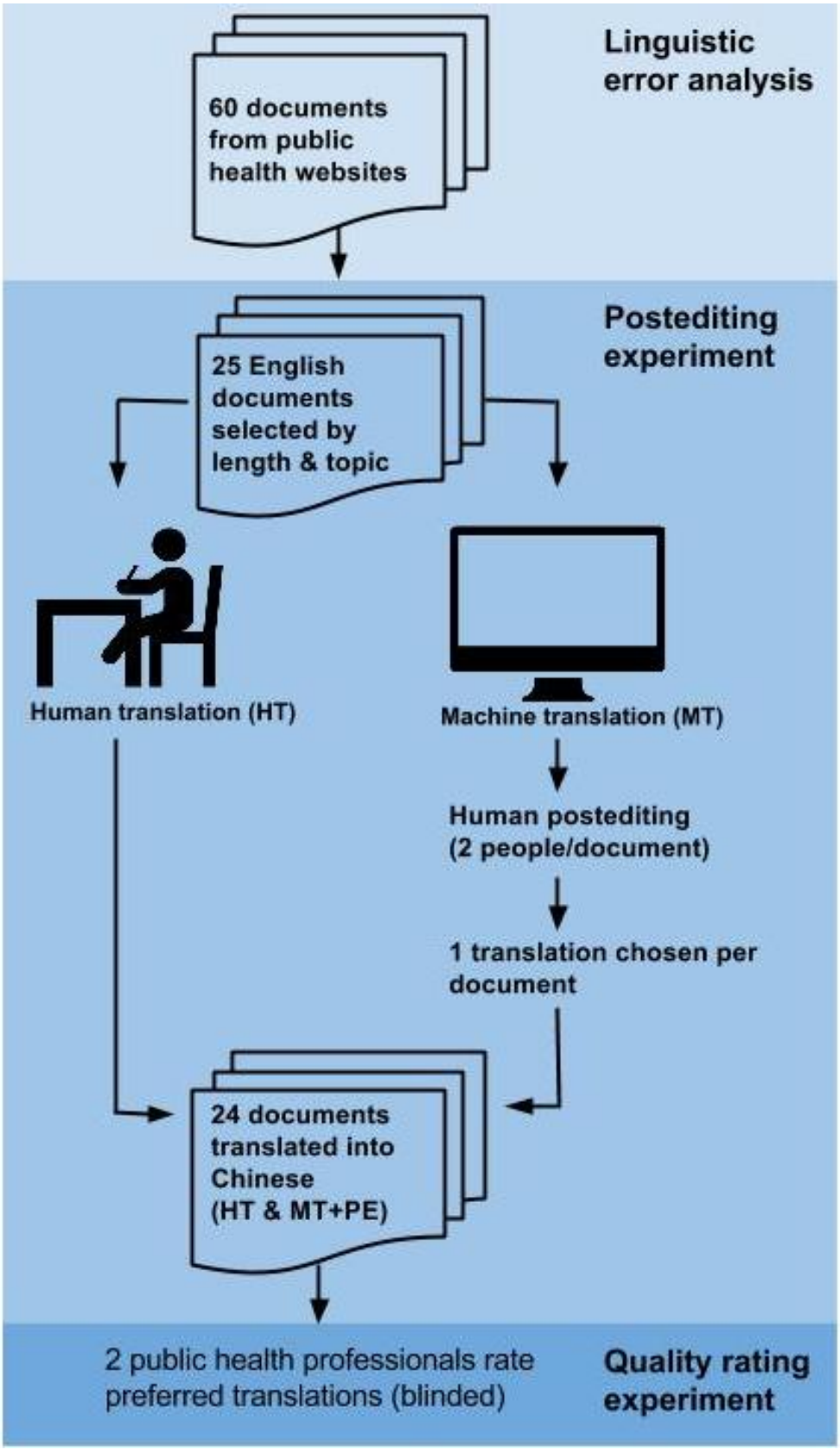
We conducted postediting experiments, similar to those conducted for the English-Spanish pair, in order to determine the feasibility (accuracy and efficiency) of using MT+PE for translating public health documents from English to Traditional Chinese. We investigated the types of MT errors occurring in Chinese, the PE time needed to correct them, and the quality of MT+PE compared to HTs, as rated by raters fluent in both English and Traditional Chinese. In this paper, we discuss the results of these investigations and compare them to our previous experiences with the English-Spanish pair. This work contributes to our understanding of the challenges involved in applying the MT+PE approach in a public health setting.

Methods

Initial Steps

We collected 60 health promotion documents from different public health agencies in the United States that had been translated manually (HT) from English to Chinese. Translations were created using the Traditional Chinese character set, as opposed to Simplified Chinese, because this is the form known to most Chinese LEP individuals in the Pacific Northwest region. We identified the types of linguistic errors present in MT from English to Chinese and then conducted the postediting of the translated materials with participants fluent in both languages. Next, we had bilingual public health professionals and laypersons rate the quality of the human versus the MT plus postedited documents. A diagram of the study design is shown in Figure 1. A more detailed description of the specific methods for the linguistic error analysis, postediting and rating studies, and follow-up evaluation is provided below.

Figure 1. Study Design Overview.



Linguistic Error Analysis

We collected 60 health promotion documents available in English and Chinese (Traditional) from public health websites in the United States. Websites included those of the Centers for Disease Control and Prevention, New York City Department of Health and Public Health, Minnesota Department of Health, Washington State Department of Health, Department of Public Health – Los Angeles County, and Public Health – Seattle & King County. All Chinese versions of these documents had been

translated manually (HT) by health department translators or professional translation vendors. The English versions of the documents were then translated into Traditional Chinese using Google Translate. We developed a categorization scheme for MT errors, and all MTs were annotated based on this scheme by a native Chinese speaker with formal training in linguistics. Subsequently, aggregate error statistics were computed to gain insights into the most frequent error categories: word sense, word order, missing word, superfluous word, orthography/punctuation, particle error, untranslated word, pragmatic error, and other grammar error.

Postediting Experiments

For the postediting studies, we selected 25 of the 60 health documents that had been machine translated from English to Chinese using Google Translate. To ensure a wide representation of topics, we selected the documents based on the length of the English version (340-914 words) and topic area. From the memberships of local Chinese cultural organizations, 6 Chinese translators were recruited for postediting and screened for language ability and health experience. Posteditors, all native Chinese speakers, were fluent in oral and written Traditional Chinese and English, had varying levels of translation experience, and had prior experience in a health-related field (Table 1).

The 25 machine-translated documents were each corrected by at least 2 posteditors in order to permit consistency checks across posteditors and computation of average time, adequacy, and fluency ratings per document. Posteditors used a proprietary MT and postediting tool built for the purpose of this study, as described previously [7]. Each posteditor corrected between four and 21 documents representing common types of public health materials, including informational webpages, agency letters, fact sheets, and brochures. Posteditors were allowed to choose their preferred character input method. One posteditor used a pinyin keyboard called Q9, while the rest used the standard Windows OS pinyin input. The postediting tool displays three versions of the text from left to right in one window: the original English text, the MT, and the editable MT, respectively. When a posteditor clicks the editable MT field to begin editing, a timer starts. The tool saves the total editing time (minus pauses), keystrokes, and a copy of the postedited machine translation. Time and keystroke data were collected for all postedited documents. Due to a posteditor saving error, only 24 of the 25 postedited documents were put out in a readable format and therefore available for rating.

Posteditors were given written and verbal instructions to "perform all corrections necessary to ensure that the text (1) is consistent with the grammar rules of Chinese, (2) adequately represents the meaning of the English text, (3) is culturally appropriate (ie, not unintentionally funny or offensive), and (4) preserves the linguistic style of the source document." Posteditors were asked not to alter a correct, appropriate translation simply because it may not correspond to their first choice of translation. In short, they were instructed to correct only as much as necessary and to not rewrite the text. These were the same instructions used in the previous Spanish study.

After completing postediting, participants were asked to fill out a questionnaire to rate the adequacy and fluency of each MT+PE on a scale of 1-5. These rating scales are common in human evaluations of machine translation quality [10]. An adequacy of 1 indicated that none of the original meaning of the English source text was retained in the MT, while an adequacy of 5 indicated that all of the meaning was retained. A fluency rating of 1 indicated that the MT was incomprehensible, while a rating of 5 indicated flawless Chinese. The questionnaire also asked participants to describe the common translation errors they found, identify which errors were most difficult to correct, and explain which errors took the longest time to correct.

Quality Rating

Two public health professionals, blinded to the method of translation, compared the quality of the postedited documents to the quality of the HT documents from the health department websites. The quality raters were asked to rate the MT+PE against HT versions. One rater was a professional public health translator and a Department of Social and Health Services Certified Medical Interpreter at a local clinic; the other was a health researcher (Table 1). They were presented with 20 sets of documents selected from the 24 available, with each set containing an original English text, an HT version of that text, and an MT+PE version of the text. Even though one rater participated in the initial postediting study as well, she did not rate documents that she had encountered while postediting. The documents were not labeled as human- or machine-translated, and the order in which they were presented in each set was randomized. Using a questionnaire, we asked the quality raters to read each set carefully, indicate which of the translated versions they preferred, and describe why they chose that version, based on five dimensions: grammar, adequacy, word choice, cultural appropriateness, and reading level.

Table 1. Initial postediting and quality rating participants, health, and translation experience.

Participant number	Role	Health background	Translation experience
P1	Posteditor	Pharmacy student	Limited—translating at health fairs
P2	Posteditor	Social work for Chinese population, including health care support	Teaching English as a second language & translating research
P3	Posteditor and quality rater	Public health researcher	10 years of various translation experience
P4	Posteditor	Social work for Chinese population, including health care support	Translating agency and government publications for distribution to clients
P5	Posteditor	Public health student	None
P6	Quality rater (posteditor for follow-up evaluation only)	Public health translator	DSHS Certified Medical Interpreter

Follow-Up Evaluation

After analyzing the results of the quality rating study, we performed follow-up evaluations of the effects of posteditor expertise, engagement, and instructions on the quality of postedited translations. To assess whether posteditors' public health and translation expertise negatively impacted the quality rating outcome, we asked P6, a highly trained and experienced health translator, to postedit four documents. We then repeated the quality rating procedure with those documents, asking five native Chinese speakers to review them.

To test posteditor engagement and whether the instructions to edit only as necessary were problematic, we asked 3 posteditors (P2, P4, and P5) to return to edit a total of 10 more documents, this time with instructions to make as many corrections as needed to ensure the quality of the translation. We again repeated the quality rating procedure with one native Chinese speaker who has public health experience to see whether posteditors given the revised instructions would produce text equivalent to the HTs.

Results

Linguistic Error Analysis

Results from the linguistic error analysis are summarized in Table 2. The left-hand column shows the error type; the right-hand column shows the corresponding frequency of the error type, computed as the percentage of all errors annotated in the total set of 60 documents. For example, word sense errors (errors where the word meaning was translated incorrectly) constituted 40% of all annotated errors. The next most common error types involved word order (22%) and missing words (16%).

Table 2. Error categories and their distributions.

Error categories	Frequency (%)
Word sense	40
Word order	22
Missing word	16
Superfluous word	14
Other grammar error	3
Orthography/punctuation	3
Particle error	1
Untranslated word	0.03
Pragmatic error	0.01

Postediting Experiments

The proprietary postediting tool recorded the time taken to postedit each machine-translated document. We analyzed the time taken, by document and by posteditor, and examined posteditors' quality ratings of the initial MT output. A list with descriptions of the source documents is provided in Multimedia Appendix 1.

To determine and analyze the amount of time required for postediting, we calculated the number of characters per minute (CPM) for each document and then computed means and standard deviations (SDs) in CPM for each document, using posteditors' recorded times. In addition, we computed means and SDs in CPM for each posteditor (Table 3). This helped us gain insights into potential correlations between postediting time and document topic, length, etc, as well as differences between posteditors (though not all posteditors edited the same number of documents).

The mean CPM per document varied greatly, from 18.5-79.6 CPM (SD 0.03-38.7). The total mean CPM across all documents was 37.8 (SD 10.2). Thus, on average a posteditor corrected approximately 38 CPM, with a variation of around 10 CPMs.

The results did not indicate a linear relationship between document length and average postediting time. We also found no relationship between the document type and the average CPM.

On average, the posteditors rated the adequacy of the translations at 3.32 (SD 0.90), suggesting that much of the original meaning of the source text was preserved in the MT. Average fluency rating was 3.0 (SD 0.84), which corresponds to a grammar quality level of non-native Chinese. The average adequacy and fluency ratings bore no relationship to the document type or length, but varied greatly by individual posteditor. Interestingly, the posteditors who had more experience with translation and health rated the adequacy and fluency lower than did their less experienced counterparts (Table 3).

To investigate the variation in postediting speed for individuals, we calculated the average CPM for each posteditor. As shown in Table 3, the average CPM was 37.4 and the average SD for CPM per document was 15.7. We also found large individual differences in speed among posteditors [11,12]. Posteditors also varied widely in their adequacy and fluency ratings, with a trend indicating an inverse relationship between public health translation experience and ratings; the more experienced posteditors in terms of translation and public health expertise tended to rate the documents they postedited lower than those with less experience (Tables 1 and 3).

Errors described by posteditors as difficult to correct, or annoying, included word sense errors and word order errors. Some examples of the errors noted by posteditors are provided in Table 4.

Table 3. Postediting time, adequacy, and fluency ratings by posteditor.

Posteditor	Docs postedited, n	CPM, mean (SD)	Avg. adequacy	Avg. fluency
P1	9	34.2 (7.3)	4	3.2
P2	21	35.4 (16.2)	N/A	N/A
P3	4	25.8 (10.2)	3	2.5
P4	4	54.3 (40.5)	3.25	3.25
P5	11	54.0 (16.0)	3.875	3.75
P6	4	20.6 (3.7)	1.75	1.625

Table 4. Posteditor examples of top three error categories.

Error category	Quotes/examples
Word sense	“The literal meaning changes when translated into Chinese (eg, lost power/electricity is translated as lost ‘energy’)”
Word order	“‘...when...can’t...’ type of sentence doesn’t have same structure in Chinese. The order of the words change in Chinese and English in many situations”
Missing word	“Whenever there is the word ‘person’ we should mention ‘this’ or ‘that’ person, otherwise it is not clear who are we talking about in the sentence.”

Quality Rating

Unlike our previous experience with English to Spanish translations, in a blind comparison of HT and MT+PE, the quality raters selected the HT document as the preferred version for all 20 documents. Reasons given for the preference were better word order, a more professional reading level, smoother flow, more accurate translated word use, preserved meaning, and cultural appropriateness of the original English document. The reasons the rater gave for rejecting the MT+PE documents were that they did not meet the reading level of the general public, some of the sentences lost the intended meaning, the same words were not translated consistently, awkward word order, and occasionally wrong word translations and awkward word flow.

Follow-Up Evaluation

In theory, if posteditors have sufficient training, experience, and resources to perform quality postediting, MT+PE documents should be equivalent to HT documents. The feasibility of utilizing MT+PE has been repeatedly demonstrated in various previous studies for a variety of language pairs; it is also a procedure that is widely used by many commercial language service providers. In previous work with the Spanish-English language pair, we found our approach feasible even among lay users with minimal training; these conditions closely mirror the public health context, where resources for training and calibration are limited.

There are several potential reasons for the preference for the HT over the MT+PE in this study:

Differences in MT Quality

Chinese machine translations have a different relative frequency of certain error types and lower quality overall. Compared to our previous studies on English-Spanish [8,13], we found that the Chinese translations had high percentages of word order and word sense errors, which require more cognitive effort to correct [14-16]. Adequacy and fluency also had lower ratings compared to the Spanish translations: adequacy for Chinese was 3.3 compared to 4.2 for Spanish; fluency was 3.1 versus 3.7 for Spanish. It should be noted that these scores are not directly comparable since the the sets of English documents used in these two studies were not identical; however, the

differences in scores confirm the common observation in the MT community that MT for English-Chinese is less effective than for English-Spanish.

Instructions Provided to Posteditors

Posteditors might have misinterpreted the postediting instructions. Specifically, the instruction to "postedit only where necessary" and to not "rewrite" might have led them to produce fewer edits than they would under real-life circumstances. Quality raters observed that the postedited documents often contained very literal word-by-word translations that were perceived as unacceptable. In other language pairs with similar linguistic structures (like English and Spanish), more literal translations may still yield acceptable translation outputs, whereas fluent Chinese requires the translator to depart more strongly from a literal translation. Due to time and resource constraints for this study, as with prior studies, there also was no extensive training and calibration phase for the study participants. Combined with the lower quality of initial MT Chinese versions, the postediting instructions might help explain the lesser quality of the postedited Chinese translations as compared to the Spanish translations.

Linguistic Expertise of Posteditors

Although posteditors were selected for bilingual competence and familiarity with the domain of public health, they did not have to undergo initial language or translation tests to verify their editing abilities.

Engagement of Posteditors

Posteditors may not have been sufficiently engaged in the task, or they may have optimized for time rather than quality.

Different Levels of Quality Control

In the postediting, only one round of postediting was performed, followed by the quality rating task. We do not know how many iterations of editing and quality control were applied to the human-generated translations, since they were collected from different sources where the translation processes were not transparent. Our prior investigations into health department translation processes revealed that most of the public health HT documents had been translated in-house or by language service providers who conduct several rounds of postediting and review prior to making them public [7].

Additional Follow-Up

In order to ascertain the contribution of these factors to the overall results, we conducted additional follow-up studies investigating the role of posteditor expertise, instruction, and engagement.

Expertise

To assess whether posteditor expertise played a role in the translation quality, we engaged the services of a public health professional who performed translation for a large metropolitan health department in Washington State (P6). She was given the original set of instructions to correct only as much as needed and to not rewrite the text extensively. She postedited four documents, which were then given as a set and blindly rated against their original human translations by five native Chinese speakers so that each rater reviewed all four documents. Three of the 5 raters selected the human translation over the MT+PE for all four documents; 2 raters rated one of the HT and MT+PE documents as equivalent.

Instructions and Engagement

To test whether our instructions to postedit only where necessary played a role in the MT+PE ratings, we modified the instructions to emphasize quality and recruited 3 posteditors to come back for another postediting session with the new instructions. The original instructions—as adapted from the Spanish study—directed posteditors to not alter a correct translation, even if it was not their first choice; to not engage in extensive rewriting of the text; and to not spend an extended period of time looking up grammar, punctuation, or unfamiliar terminology online. The updated instructions directed posteditors to use as much time and effort as necessary to ensure a high-quality translation. The 3 returning posteditors corrected a total of 10 documents, which were then blindly rated by a quality rater with language and public health expertise. As anticipated, posteditors took longer to produce the MT+PE translations with the updated instructions: P2's average speed dropped from 35.38 CPM to 23.43 CPM, P4's fell from 54.33 to 17.46 CPM, and P5's decreased from 53.96 to 19.69 CPM. The rater chose the manual human translations for 6/10 documents, while rating four as equivalent—a notable improvement over the original instructions.

Discussion

Principal Findings

Although our prior research on English to Spanish translation indicated that MT+PE could produce translations equivalent in quality for less time and cost, our current study on the English-Chinese language pair showed that maintaining quality through postediting was more problematic. Translation between English and Chinese presents a challenge due to very divergent syntactic structures (eg, topic-comment structure in Chinese vs subject-verb-object structure in English), frequent dropping of pronouns in Chinese, higher degree of morphology in English, and other linguistic differences. Compared to a language pair like English and Spanish, SMT for English and Chinese generally tends to produce lower-quality results (eg, the results obtained in benchmark evaluations for different language pairs conducted by the US National Institute of Standards and Technology [17].

Strengths and Limitations

Although, theoretically, professional translators with sufficient training and time should be able to produce an equivalent product through postediting MTs, even with instructions to take the time to provide the best quality translation, the final postedited translations still contained obvious errors that led the quality raters to prefer HTs in most cases. Experienced translators who performed the translations rated the adequacy and fluency of MT+PE lower in general than their less experienced counterparts and commented that for many machine-translated sentences it would be easier to start with the English version than to correct the MT version. However,

it should be noted that our prior evaluation of health department translation processes found that HT documents undergo multiple editing cycles to ensure translation quality and cultural appropriateness. In the studies reported on here, the machine-translated documents underwent only one round of postediting. It is likely that with additional rounds of editing the MT+PE product would be further improved.

Another possible limitation of our study is the use of a single translation engine, Google Translate. However, most SMT systems are based on the same set of underlying statistical models, suggesting that the types and relative frequencies of translation errors would not have been significantly different had a different SMT system been used.

Additional work is needed to improve the quality of MT from English to Chinese. Word sense and word order errors require the most attention for improvement. Our team is currently working to improve these errors. In addition, particular care must be taken in selecting posteditors, documents, and machine translation engines, and in designing postediting instructions and quality control processes.

Conclusion

In the United States, Chinese is the second most common language spoken by LEP individuals and the single most common character language used. However, due to the resources and time involved in human translation, health departments currently offer few health promotion materials in Chinese. Our investigation into the use of MT+PE to produce translations indicates that using the methods that worked for English to Spanish translations was not as effective with translation from English to Chinese. Multiple factors, including quality of MT and expertise of posteditors, may have contributed to these results. Our preliminary follow-up studies suggest that reducing word sense errors and word order errors would improve English to Chinese MTs, while additional training and expertise of bilingual posteditors may be needed in order to successfully apply online MT technology to public health practice. We are performing additional studies to determine how best to improve translation from English to Chinese in order to ensure quality translation at a low cost.

Acknowledgments

The research reported here was supported by the National Library of Medicine of the National Institutes of Health (NIH) under award number R0110432704. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH. The images used in Figure 1 were created by Hadi Davodpour, Edward Boatman, and Lauren Manninen for the Noun Project. We would also like to thank Beryl Schulman and Julie Loughran for reviewing this manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Study source documents and postediting times.

[PDF File (Adobe PDF File), 41KB - publichealth_v1i2e17_app1.pdf]

References

1. Turner A, Capurro D, Kirchhoff K. 3rd Annual Health Literacy Research Conference. Chicago, IL: Health Literacy Research Conference; 2011. The availability of translated public health materials for limited English proficiency populations in Washington State URL: http://www.bumc.bu.edu/healthliteracyconference/files/2011/07/Poster-Abstracts-Packet.pdf [accessed 2015-05-19] [WebCite Cache ID 6Ye9EvBZk]
2. Raynor EM. Factors Affecting Care in Non-English-Speaking Patients and Families. Clin Pediatr (Phila) 2015 May 11. [doi: 10.1177/0009922815586052] [Medline: 25963177]
3. Ponce NA, Hays RD, Cunningham WE. Linguistic disparities in health care access and health status among older adults. J Gen Intern Med 2006 Jul;21(7):786-791 [FREE Full text] [doi: 10.1111/j.1525-1497.2006.00491.x] [Medline: 16808783]
4. Sentell TL, Tsoh JY, Davis T, Davis J, Braun KL. Low health literacy and cancer screening among Chinese Americans in California: a cross sectional analysis. BMJ Open 2015;5:1-9.
5. Aranberri N, Labaka G, Diaz de Ilarraza A, Sarasola K. Comparison of Post-Editing Productivity between Professional Translators and Lay Users. In: Third Workshop on Post-editing Technology and Practice. 2014 Oct Presented at: Third Workshop on Post-editing Technology and Practice; October 2014; Vancouver (BC), Canada p. 20-33 URL: http://www.amtaweb.org/AMTA2014Proceedings/AMTA2014Proceedings_PEWorkshop_final.pdf
6. Koehn P, Germann U. The impact of machine translation quality on human post-editing. In: Workshop on Humans and Computer-Assisted Translation. 2014 Presented at: Workshop on Humans and Computer-assisted Translation; 2014; Gothenburg, Sweden p. 38-46 URL: http://www.aclweb.org/anthology/W14-0307.pdf

7. Turner AM, Bergman M, Brownstein M, Cole K, Kirchhoff K. A comparison of human and machine translation of health promotion materials for public health practice: time, costs, and quality. J Public Health Manag Pract 2014;20(5):523-529. [doi: 10.1097/PHH.0b013e3182a95c87] [Medline: 24084391]
8. Kirchhoff K, Turner AM, Axelrod A, Saavedra F. Application of statistical machine translation to public health information: a feasibility study. J Am Med Inform Assoc 2011;18(4):473-478 [FREE Full text] [doi: 10.1136/amiajnl-2011-000176] [Medline: 21498805]
9. Pandya C, Batalova J, McHugh M. Limited English Proficient Individuals in the United States: Number, Share, Growth, Linguistic Diversity.: Migration Policy Institute; 2011. URL: http://www.immigrationresearch-info.org/report/migration-policy-institute/limited-english-proficient-individuals-united-states-number-share- [accessed 2015-05-17] [WebCite Cache ID 6Ye8R4lAY]
10. Linguistic Data Consortium. Linguistic Data Annotation Specification: Assessment of Fluency and Adequacy in Translations Revision 1.5. 2005 Jan 25. URL: https://www.ldc.upenn.edu/collaborations/past-projects [accessed 2015-05-18] [WebCite Cache ID 6Ye9VQH0J]
11. Guerberof A. Machine Translation Summit XII. 2009 Aug. Productivity and quality in MT post-editing URL: http://www.mt-archive.info/MTS-2009-Guerberof.pdf [accessed 2015-05-18] [WebCite Cache ID 6Ye9wRXwj]
12. Guerberof A. Correlations between productivity and quality when post-editing in a professional context. Machine Translation 2014 Nov 20;28(3-4):165-186. [doi: 10.1007/s10590-014-9155-y]
13. Kirchhoff K, Capurro D, Turner AM. A Conjoint Analysis Framework for Evaluating User Preferences in Machine Translation. Mach Transl 2014 Mar 1;28(1):1-17 [FREE Full text] [doi: 10.1007/s10590-013-9140-x] [Medline: 24683295]
14. Temnikova I. Cognitive Evaluation Approach for a Controlled Language Post-Editing Experiment. In: Proceedings of the Seventh International Conference on Language Resources and Evaluation. 2010 May Presented at: Seventh International Conference on Language Resources and Evaluation; May 2010; Valletta, Malta p. 3485-3490 URL: http://www.lrec-conf.org/proceedings/lrec2010/pdf/437_Paper.pdf
15. Lacruz I, Denkowski M, Lavie A. Cognitive Demand and Cognitive Effort in Post-Editing. In: Third Workshop on Post-Editing Technology and Practice. 2014 Oct Presented at: The Third Workshop on Post-Editing Technology and Practice; October 2014; Vancouver (BC), Canada p. 73-84 URL: http://www.amtaweb.org/AMTA2014Proceedings/AMTA2014Proceedings_PEWorkshop_final.pdf
16. Koponen M, Aziz W, Ramos L, Specia L. Post-Editing Time as a Measure of Cognitive Effort. In: Workshop on Post-Editing Technology and Practice. 2012 Oct Presented at: Tenth Biennial Conference of the Association for Machine Translation of the Americas; October 2012; San Diego, California URL: http://amta2012.amtaweb.org/AMTA2012Files/html/13/13_paper.pdf
17. Koehn P. Options, Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the ACL. 2010 Jun Presented at: 2010 Annual Conference of the North American Chapter of the ACL; June 2010; Los Angeles, CA p. 537-545 URL: http://www.aclweb.org/anthology/N10-1078

Abbreviations

CPM: characters per minute
HT: human translation
LEP: limited English proficiency
MT: machine translation
NIH: National Institutes of Health
PE: postediting
SMT: statistical machine translation

Edited by G Eysenbach; submitted 29.05.15; peer-reviewed by D Capurro, Y Kim, B Massoudi; comments to author 08.07.15; revised version received 18.08.15; accepted 07.10.15; published 17.11.15.

Please cite as:
Turner AM, Dew KN, Desai L, Martin N, Kirchhoff K
Machine Translation of Public Health Materials From English to Chinese: A Feasibility Study
JMIR Public Health Surveill 2015;1(2):e17
URL: http://publichealth.jmir.org/2015/2/e17/
doi:10.2196/publichealth.4779
PMID:27227135

Original Paper

Patient-Reported Outcomes and Total Health Care Expenditure in Prediction of Patient Satisfaction: Results From a National Study

Man Hung[1,2,3,4,5,6], PhD; Weiping Zhang[4], PhD; Wei Chen[5], PhD; Jerry Bounsanga[1]; Christine Cheng[1]; Jeremy D Franklin[6], PhD; Anthony B Crum[1], BS; Maren W Voss[1], MS; Shirley D Hon[1]

[1]Department of Orthopaedics, University of Utah, Salt Lake City, UT, United States
[2]Division of Public Health, University of Utah, Salt Lake City, UT, United States
[3]School of Business, University of Utah, Salt Lake City, UT, United States
[4]School of Medicine, University of Utah, Salt Lake City, UT, United States
[5]Division of Epidemiology, University of Utah, Salt Lake City, UT, United States
[6]College of Education, University of Utah, Salt Lake City, UT, United States

Corresponding Author:
Man Hung, PhD
Department of Orthopaedics
University of Utah
590 Wakara Way
Salt Lake City, UT, 84108
United States
Phone: 1 801 587 5372
Fax: 1 801 587 5411
Email: man.hung@hsc.utah.edu

Abstract

Background: Health care quality is often linked to patient satisfaction. Yet, there is a lack of national studies examining the relationship between patient satisfaction, patient-reported outcomes, and medical expenditure.

Objective: The aim of this study is to examine the contribution of physical health, mental health, general health, and total health care expenditures to patient satisfaction using a longitudinal, nationally representative sample.

Methods: Using data from the 2010-2011 Medical Expenditure Panel Survey, analyses were conducted to predict patient satisfaction from patient-reported outcomes and total health care expenditures. The study sample consisted of adult participants (N=10,157), with sampling weights representative of 233.26 million people in the United States.

Results: The results indicated that patient-reported outcomes and total health care expenditure were associated with patient satisfaction such that higher physical and mental function, higher general health status, and higher total health care expenditure were associated with higher patient satisfaction.

Conclusions: We found that patient-reported outcomes and total health care expenditure had a significant relationship with patient satisfaction. As more emphasis is placed on health care value and quality, this area of research will become increasingly needed and critical questions should be asked about what we value in health care and whether we can find a balance between patient satisfaction, outcomes, and expenditures. Future research should apply big data analytics to investigate whether there is a differential effect of patient-reported outcomes and medical expenditures on patient satisfaction across different medical specialties.

(JMIR Public Health Surveill 2015;1(2):e13) doi:10.2196/publichealth.4360

KEYWORDS

health care quality; value; expenditure; cost; medical outcomes; patient satisfaction; Medical Expenditure Panel Survey; patient-reported outcomes; Affordable Care Act; big data analytics

Introduction

Value-based health care has become a buzzword for health care reforms across the globe. In the United States, the Affordable Care Act has placed huge emphasis on health care value and quality [1]. Although it has yet to be clearly defined, health care quality is often linked to patient satisfaction [2-4]. Beginning on October 1, 2012, the Center for Medicare & Medicaid Services (CMS) has tied Medicare reimbursements with patient satisfaction, as measured by the Hospital Consumer Assessment of Healthcare Providers and Systems (CAHPS) survey. The CAHPS survey measures the following 5 aspects of patient satisfaction or perception of health care experiences: (1) access to care, (2) provider communication, (3) coordination of care, (4) shared decision-making, and (5) the office staff. Additionally, it contains a global rating item on patient satisfaction. The results of the patient satisfaction survey are estimated to put a hospital at risk of US $500,000 to US $850,000 on average for Medicare reimbursement [5]. The stakes are high with patient satisfaction, but how can we improve patient satisfaction and what are the predictors of patient satisfaction?

One question is "what constitutes value in health care?" Porter posited that health care value is linked to patient-reported outcomes (PROs) and expenditures [6]. PROs are patients' self-reported health status based on their perceptions of their health conditions. PROs, in addition to information from clinical assessments, have become a critical component of medical assessment by directly receiving information from the patient's perspective [7]. Some argue that understanding PROs is critical to understanding patient satisfaction [8]. PRO instruments, such as SF-12 v2, can measure various outcomes including physical health, mental health, and general health.

While PROs have become increasingly important in assessing health care value, the second part of the value equation is cost or expenditures in health care. Prior research has showed mixed findings between expenditures and patient satisfaction [9-12]. Fenton et al [11] found that higher inpatient utilization, lower emergency department utilization, higher total medical expenditures, higher prescription drug expenditures, and fewer emergency visits are associated with higher patient satisfaction. However, a review of the literature found no consistent relationship and the majority of studies only found a small association between expenditures and satisfaction [10].

Research has suggested that a number of sociodemographic characteristics correlate with patient satisfaction, but there is not a strong consensus [13]. Several studies demonstrated that older patients, patients with better functional status, and clearer communication from doctors are related to greater patient satisfaction [9,11,14-16]. Rahmqvist and Bara [17] also found that older patients were more satisfied than younger patients [17], yet critiques suggested that the magnitude of age as a predictor of satisfaction is small [18]. Additionally, greater patient satisfaction has been associated with lower levels of education and income [17-19]. However, Hall and Dornan [18] found no relationship among income, ethnicity, sex, or family size with patient satisfaction.

Although sociodemographics and other nonmodifiable characteristics may influence patient satisfaction, they are not very useful for implementing changes in health care. Modifiable characteristics such as patient outcomes and health care expenditures are more useful for attaining actionable changes and reform in health care, yet there is a shortage of research in this area. There is also a lack of longitudinal, national studies examining the relationship between PROs, medical expenditures, and patient satisfaction. This study aims to fill this gap by investigating this relationship using a longitudinal, nationally representative sample, adjusting for various sociodemographics and health-related characteristics.

Methods

Study Design

Overview

Data from the longitudinal study of adult respondents to the 2010-2011 longitudinal panel Medical Expenditures Panel Survey (MEPS) [20] served as a basis for the comprehensive assessment of health care value and health care quality in this study. The MEPS is a nationally representative survey, sponsored by the Agency for Healthcare Research and Quality, which measures access, use, and cost of health care services. The survey consists of 3 major components, namely, (1) the household, (2) the medical provider, and (3) insurance. The household component samples were drawn from the respondents to the National Health Interview Survey by the National Center for Health Statistics. This study utilized data from the household component and included respondents aged 18 years or older. A list of all the variables used in this study is shown in Multimedia Appendix 1, and Multimedia Appendix 2 contains all of the covariates included in the analyses. Capitalizing on MEPS's longitudinal panel survey design, we sought to predict patient satisfaction from PROs and total health care expenditures, controlling for demographics, prior health status, and clinical characteristics.

Medical Expenditures

The MEPS collects data on various categories of medical expenditures such as prescription drugs, emergency patient visits, and inpatient hospital stays. The variable total expenditure is an aggregate of medical expenditures in various categories. In this study, we used total health care expenditures (EXP) from Year 1 (2010) as an independent variable.

Patient-Reported Outcomes

The MEPS also contains PROs data. Specifically, it contains data from the SF-12v2, which measures patients' self-reported functional health status and well-being. The SF-12v2's physical health component score (PCS), the mental health component score (MCS), and the general health perceptions (GH) score from Year 1 were also included as independent variables in this study. The PCS and MCS possible scores range from 0 (worst health) to 100 (best health). To put GH in the same direction as the PCS and MCS scoring, we reverse coded the original GH such that the possible scores range from 1 (worst health) to 5 (best health).

Patient Satisfaction

We used the CAHPS's single-item global rating of satisfaction in Year 2 (2011) as a dependent variable. This global patient satisfaction item (SAT) reflects patients' rating of their health care from all physicians and health care providers in the last 12 months when the patients were taking the survey. It is not a recall of a certain visit at a specific time (eg, at 3 month or 5 months ago), rather it represents a patient's average experience of all health care encounters within the last 12 months. The possible scores range from 0 (worst health care possible) to 10 (best health care possible). The survey was administered using computer-assisted personal interviewing technology.

Covariates

Based on a literature review, we identified a list of potential confounders from the MEPS data to be the covariates and adjusted for them when investigating the contribution of PROs and total health care expenditure to patient satisfaction. The list of potential confounders is included in Multimedia Appendix 2, which contains patients' age, sex, education, race, ethnicity, income, insurance coverage, provider characteristics, prior health status, and different clinical conditions. The covariates selected for the regression analyses were confounders that showed significant associations with patient satisfaction.

Data Analysis

The MEPS utilized a multistage, probability clustering sample design that enabled comprehensive examination of the US population. The sampling weight, stratification, clustering, multiple stages of selection, and disproportionate sampling from the MEPS were taken into account in the analyses so that the findings reported represented the entire US population. Sampling weight took into account the differential probability of sample selection and adjusted for nonresponses and missing data. Descriptive statistics were conducted to examine SAT, PCS, MCS, GH scores, EXP, and different service categories of medical expenditures across clinical conditions. We also performed descriptive statistics on all potential confounders and flagged those that had significant associations with patient satisfaction to be included as covariates in subsequent regression analyses.

To investigate the contribution of PCS, MCS, GH, and EXP on patient satisfaction, we conducted logistic regressions with adjustment of covariates and reported the odds ratio with associated 95% CI. We recoded the PROs, satisfaction, and expenditure variables into low/high SAT, PCS, MCS, GH, and EXP prior to running the logistic regressions. All statistical tests were two sided, were set at an alpha level of .05, and were conducted using SAS 9.3. Institutional review board (IRB) and/or ethics committee approval was not required as the MEPS data are freely available to the general public online.

Results

Demographics

The entire study sample consisted of adult participants (N=10,157), which represented 233.26 million adults (aged ≥ 18 years) in the United States. The mean age was 47 years (SE 0.28, range 18-85). Approximately 52% (121 million/233.26 million) were female and the majority were white (81%, 188 million/233.26 million) and black (12%, 27 million/233.26 million). About 15% (34 million/233.26 million) were Hispanic, approximately 34% (79 million/233.26 million) were unemployed, and nearly 12% (27 million/233.26 million) spoke non-English at home (Table 1).

Table 1. Demographics of the sample from MEPS (weighted N=233.26 million; unweighted N=10,157).

Variable		Weighted n (millions)	%	Unweighted n (millions)	%
Sex					
	Male	113	48.3	2385	72.3
	Female	121	51.7	916	27.7
Race					
	White	188	80.7	7060	69.5
	Black	27	11.7	1992	19.6
	American Indian/Alaskan native	2	0.7	84	0.8
	Asian	12	5.1	805	7.9
	Native Hawaiian/Pacific Islander	2	0.7	72	0.7
	Multiple races	3	1.1	144	1.4
Ethnicity					
	Hispanic	34	14.5	2475	24.4
	Black (not Hispanic or another race)	27	11.4	1950	19.2
	Asian (not Hispanic or another race)	12	5.0	794	7.8
	Other race (not Hispanic or another race)	161	69.1	4938	48.6
Hispanic					
	Hispanic	34	14.5	2475	24.4
	Not Hispanic	200	85.5	7682	75.6
Employment status					
	Employed	154	66.2	6465	63.9
	Not employed	79	33.8	3653	36.1
Marital status					
	Married	124	53.4	5172	50.9
	Widowed	15	6.3	612	6.0
	Divorced	26	11.3	1173	11.5
	Separated	5	2.3	312	3.1
	Never married	62	26.8	2888	28.4
Currently smoke					
	Yes	39	18.3	1732	18.5
	No	176	81.7	7606	81.5
Language spoken at home					
	English	205	88.3	8009	79.2
	Spanish	19	8.1	1555	15.4
	Another language	8	3.6	542	5.4
Highest level of education					
	High school or less with no degree	31	13.4	1986	20.0
	High-school graduate or GED	69	29.9	3239	32.5
	Associates degree, beyond college, but no degree	62	26.9	2359	23.7

Variable		Weighted n (millions)	Weighted %	Unweighted n (millions)	Unweighted %
	Bachelor's degree	45	19.4	1596	16.0
	Master's, PhD, or professional degree	24	10.4	777	7.8

This nationally representative adult population had a number of health conditions. About 32% (74.97 million/233.09 million) had high blood pressure and 5% (12.31 million/233.11 million) had coronary heart disease. Approximately 4% (8.52 million/233.20 million) had previously experienced a stroke. Almost 30% (70.59 million/233.07 million) had high cholesterol and 24% (56.34 million/233.18 million) had arthritis. There were 10% (23.48 million/233.20 million) diagnosed with cancer.

The sample population visited a variety of providers; however, most visited a general/family practice physician (71%, 60.69 million/86.00 million) and internal medicine physicians (20%, 17.10 million/86.00 million) within the last 12 months. About 68% (101.57 million/150.09 million) stated that their providers always showed respect for their treatment, whereas 3% (3.89 million/150.09 million) stated that their providers never showed respect. Almost 96% (161.35 million/168.25 million) indicated that their providers explained treatment options but 4% (6.90 million/168.25 million) did not. Approximately 67% (156.18 million/233.26 million) had private insurance, 18% (41.31 million/233.26 million) had public health insurance, whereas over 15% (35.77 million/233.26 million) were uninsured. A detailed breakdown of the number of responses for each variable is not provided here but can be obtained by contacting the authors.

Medical Expenditures

The 2010 annual total health care expenditure in the US was over US $1.11 trillion and the average total health care expenditure per adult was US $4752.39 (Table 2). The total medical expenditures for emergency visits, hospital stay, dental care, and prescription drugs were US $7.17 billion, US $41.21 billion, US $62.87 billion, and US $252.45 billion, respectively. The US medical expenditures, by major service categories, by disease conditions are displayed in Table 2. Table 3 presents the descriptive statistics for medical expenditure, outcomes, and patient satisfaction.

Table 2. US medical expenditure by service area, by condition, per adult in 2010 (in US $).

		Median	Mean	SE	Minimum	Maximum
Emergency visits						
	High blood pressure	0	38.19	3.92	0	3406
	Coronary heart disease	0	60.95	11.80	0	1576
	Angina	0	77.89	16.52	0	900
	Stroke	0	56.73	11.30	0	1092
	Emphysema	0	60.23	14.10	0	701
	High cholesterol	0	36.29	4.89	0	4570
	Cancer	0	34.52	5.72	0	936
	Arthritis	0	42.72	5.41	0	4570
	Asthma	0	48.23	9.12	0	4570
	Hysterectomy	0	52.49	9.38	0	5424
	Heart attack	0	56.23	11.52	0	732
Hospital stay						
	High blood pressure	0	233.94	29.24	0	28,798
	Coronary heart disease	0	498.17	115.31	0	17,792
	Angina	0	674.49	190.66	0	17,792
	Stroke	0	480.89	123.35	0	17,792
	Emphysema	0	381.02	96.18	0	5155
	High cholesterol	0	228.66	27.37	0	17,792
	Cancer	0	229.43	41.66	0	8923
	Arthritis	0	227.38	27.28	0	17,792
	Asthma	0	183.19	33.72	0	9289
	Hysterectomy	0	214.13	25.81	0	7550
	Heart attack	0	572.91	178.98	0	17,792
Dental care						
	High blood pressure	0	235.39	20.11	0	15,692
	Coronary heart disease	0	252.75	68.19	0	10,289
	Angina	0	198.33	57.69	0	3514
	Stroke	0	180.72	38.08	0	3514
	Emphysema	0	250.03	75.47	0	5998
	High cholesterol	0	273.77	22.51	0	15,692
	Cancer	0	291.01	31.46	0	6192
	Arthritis	0	259.89	21.74	0	15,692
	Asthma	0	203.07	23.56	0	5998
	Hysterectomy	0	266.83	23.91	0	9674
	Heart attack	0	188.15	58.59	0	5998
Prescription drugs						
	High blood pressure	711.50	1934.43	82.76	0	50,667
	Coronary heart disease	2030.00	3209.79	246.04	0	23,355
	Angina	2080.50	3269.23	329.06	0	17,576
	Stroke	1459.00	2902.59	326.25	0	40,940
	Emphysema	2150.00	3292.43	334.82	0	15,071

		Median	Mean	SE	Minimum	Maximum
	High cholesterol	779.00	1985.17	85.30	0	40,940
	Cancer	691.00	2153.63	180.58	0	37,509
	Arthritis	914.50	2178.59	100.20	0	50,667
	Asthma	571.50	1863.06	130.80	0	31,763
	Hysterectomy	826.50	1826.24	104.19	0	40,940
	Heart attack	1858.00	3141.72	326.99	0	23,355

Table 3. Descriptive statistics for medical expenditure, outcomes, and patient satisfaction.

Variables	Mean (SE)	95% CI
Physical health component score (PCS)	49.28 (0.16)	48.97-49.59
Mental health component score (MCS)	50.98 (0.14)	50.7-51.25
General health (GH)	3.55 (0.02)	3.52-3.58
Total health care expenditure (EXP), US $	4752.39 (145.43)	4465.61-5039.17
Patient satisfaction (SAT)	8.30 (0.03)	8.24-8.36

Patient-Reported Outcomes

The mean score for PCS, MCS, and GH was 49.28 (SE 0.16), 50.98 (SE 0.14), and 3.55 (SE 0.02), respectively. These scores were significantly correlated with each other (PCS with MCS, ρ=.16, *P*<.001; PCS with GH, ρ=.96, *P*<.001; MCS with GH, ρ=.62, *P*<.001).

Patient Satisfaction

As measured by CAHPS's global rating, the mean patient satisfaction score across the national adult sample was 8.30 (SE 0.03). The following were some of the variables that were significantly associated with patient satisfaction: age (ρ=.25, *P*<.001), currently smoking (χ^2_1=29.7, *P*<.001), provider showed respect for treatment (χ^2_3=75.5, *P*<.001), and provider explained option to person (χ^2_1=21.0, *P*<.001). The complete list of all of the variables that were found to be associated with patient satisfaction is provided in Multimedia Appendix 2.

Patient-Reported Outcomes, Medical Expenditures, and Patient Satisfaction

Total health care expenditure was negatively related to PROs; individuals who had lower physical function (ρ=−.43, *P*<.001), lower mental function (ρ=−.11, *P*<.001), and lower general health (ρ=−.51, *P*<.001) had higher total health care expenditures. After adjusting for covariates, we found that PROs and total health care expenditure were highly related to patient satisfaction (Table 4). Higher physical function, higher mental function, higher general health status, and higher total health care expenditure were associated with higher patient satisfaction. The odds of those who had high GH being satisfied were 6 times greater than those who had low GH (adjusted OR 5.98, 95% CI 2.95-12.12). High GH was defined as those who had excellent GH and low GH was defined as those who had poor or fair GH. There was more than a 2-fold difference in patient satisfaction between those who had high and low PCS (adjusted OR 2.54, 95% CI 1.36-4.72) and a 2-fold difference between those who had high and low MCS (adjusted OR 1.95, 95% CI 1.20-3.15). Those who had high EXP being satisfied were 3 times greater than those who had low EXP (adjusted OR 3.20, 95% CI 1.47-6.98). High PCS, MCS, and EXP were defined by those respondents who had ≥75 percentile of the scores, whereas low PCS, MCS, and EXP were defined by those with <25 percentile scores.

Table 4. Prediction of patient satisfaction.

Independent variables	n	Adjusted odds ratio[a]	95% CI
Physical health component score (PCS)	4782	2.54	1.36-4.72
Mental health component score (MCS)	4728	1.95	1.20-3.15
General health (GH)	3251	5.98	2.95-12.12
Total health care expenditure (EXP)	5082	3.20	1.47-6.98

[a]Covariates are listed in Multimedia Appendix 2.

Discussion

Principal Findings

The purpose of this study was to investigate the contribution of physical health, mental health, general health, and total health care expenditures to patient satisfaction. We found that higher scores of physical health, mental health, and general health were related to higher scores of patient satisfaction. Consistent with previous research, we found that some factors such as age and provider interactions could impact patient satisfaction. Additionally, total health care expenditure was associated with patient satisfaction. Taken together, findings from this study can help providers, payers, policy makers, and the general public better understand the relationship between PROs, health care expenditures, and satisfaction.

General health demonstrated the strongest relationship with patient satisfaction. After accounting for demographics and various factors, we see that greater patient satisfaction is directly related to greater physical and mental health, although the relationship is stronger with physical health. The argument that the relation between higher patient satisfaction and better health is an artifact of the tendency of healthy patients to be satisfied is not new [21]. This stresses the importance of controlling for prior health status as we did in this study and with the inclusion of prior clinical conditions into the model. Previous research found that a functional measure of health predicted greater satisfaction levels immediately following the medical treatment [9]. Differing research results regarding the impact of patient satisfaction on health have been partially attributed to differences in how risk or health have been assessed as well as the time frame as to when the satisfaction scores were gathered [22]. It may be that individuals who are in poor health have to visit a physician more often, increasing the likelihood of a bad or unsatisfactory visit. Prior experience with health care, along with age and mental status variables had been shown to impact patient expectations [23]. Our study emphasizes the importance of controlling for prior health status, prior health care experience, nonmodifiable characteristics, and patient expectations when assessing the effect of patient satisfaction.

Mental health often goes unnoticed or it is dismissed as someone having a bad day. Both identifying mental health issues and addressing problems require training that goes beyond that of a general physician. As a result, a patient or a provider often needs to raise a concern about mental health. A stigma still exists about discussing and treating mental health issues, which may lead patients to be less willing to discuss mental health conditions. Seeing a mental health specialist adds an additional cost for patients that they may want to avoid. Finally, patients might not even notice that they have mental health issues. These may be reasons why mental health is a weaker predictor of patient satisfaction than physical health.

A surprising finding was that total health care expenditures also had a significant relation to patient satisfaction. In a recent review of health care quality, only one third of the 61 studies reviewed found a positive association between higher spending and better health care quality [10], but did not address patient satisfaction directly as a measure of quality. While our findings add to the body of research that suggests an association between spending and patient satisfaction [11], it contradicts other research that states there is no association [9,24]. It is possible that initiatives that improve PROs and control cost at the same time contribute to patient satisfaction. Instead of raising health care expenditures, health policy makers and providers should attend to modifiable characteristics that are within their control such as providers showing respect and providers explaining treatment options to enhance PROs and thus, patient satisfaction.

Limitations

We were limited by the measurement tools in the MEPS, thus limiting the data we could use to measure physical, mental, and general health. Newer instruments have been developed using advanced techniques and these instruments may have better psychometric properties than instruments included in MEPS. The MEPS was administered using computer-assisted personal interviewing technology and the results obtained from this mode of survey administration may be different from those obtained from other modes. Care needs to be taken when doing cross-comparisons of results from different modes in different studies. Furthermore, the definition of health care expenditures reflects the CMS's definition but might not represent a complete assessment of medical expenditures. We were not able to look at diseases that were not included in MEPS. Even though we had adjusted for nonresponses and missing data in the sample, such adjustment could only be made to a certain degree. The adjustment might not be adequate in the event that nonrespondents differed from respondents. It is well-known that there could be a larger share of nonrespondents having severe illness, impairment, and dealing with poor social economic conditions on a daily basis. Additionally, the majority of the population is white, but within the past few years, the United States has seen a growth of racial and ethnic diversity. Therefore, our findings might not represent underrepresented minorities at this moment in time.

Finally, the concept of patient satisfaction lacks a clear connotation and definition in the literature. It has sometimes been considered as a component of consumer marketing [6] that measures important customer service qualities within the realm of health care. These qualities include effective provider communication, support from physicians, waiting time, fulfillment of patient requests, staff integrity, and shared decision making, and represent multiple dimensions of patient satisfaction. Together, these dimensions of patient satisfaction constitute a patient's overall health care experience. It is this overall patient health care experience that this study used to define patient satisfaction. It is likely that we were not accounting for everything that comprises patient satisfaction. In the future it would be beneficial to collect different aspects of patient satisfaction data and examine them under a different lens.

Future Research Directions

Future research should consider other modifiable characteristics that may influence patient satisfaction so that changes can be enacted. Often, underrepresented minorities have different experiences with health care than European Americans. Therefore, investigating their perspectives may illuminate group

differences. Finally, future research should apply business intelligence or big data analytics to investigate whether there is a differential effect of PROs and medical expenditures on patient satisfaction across different medical areas.

Conclusions

We found that PROs and total health care expenditure had a strong relationship with patient satisfaction. As more emphasis is placed on health care value and quality, this area of research will become increasingly needed. Critical questions should be asked about what we value in health care and whether we can find a balance between patient satisfaction, outcomes, and expenditures. These questions need to be asked to policy makers, physicians, patients, insurers, and the general public.

Conflicts of Interest

None declared.

Multimedia Appendix 1

List of all MEPS variables examined in this study.

[PDF File (Adobe PDF File), 144KB - publichealth_v1i2e13_app1.pdf]

Multimedia Appendix 2

Association between patient satisfaction and potential confounders. Variables are rank listed with the strongest relations to patient satisfaction at the top and the lowest relations at the bottom.

[PDF File (Adobe PDF File), 313KB - publichealth_v1i2e13_app2.pdf]

References

1. Reineck LA, Kahn JM. Quality measurement in the Affordable Care Act: a reaffirmed commitment to value in health care. Am J Respir Crit Care Med 2013 May 15;187(10):1038-1039. [doi: 10.1164/rccm.201302-0404ED] [Medline: 23675709]
2. Sequist TD, Schneider EC, Anastario M, Odigie EG, Marshall R, Rogers WH, et al. Quality monitoring of physicians: linking patients' experiences of care to clinical quality and outcomes. J Gen Intern Med 2008 Nov;23(11):1784-1790 [FREE Full text] [doi: 10.1007/s11606-008-0760-4] [Medline: 18752026]
3. Kupfer JM, Bond EU. Patient satisfaction and patient-centered care: necessary but not equal. JAMA 2012 Jul 11;308(2):139-140. [doi: 10.1001/jama.2012.7381] [Medline: 22782413]
4. Glickman SW, Boulding W, Manary M, Staelin R, Roe MT, Wolosin RJ, et al. Patient satisfaction and its relationship with clinical quality and inpatient mortality in acute myocardial infarction. Circ Cardiovasc Qual Outcomes 2010 Mar;3(2):188-195 [FREE Full text] [doi: 10.1161/CIRCOUTCOMES.109.900597] [Medline: 20179265]
5. Buhlman N, Matthes N. White Papers. South Bend, IN: Press Ganey; 2011 May 01. The time to prepare for value-based purchasing is now URL: http://pressganey.com/Documents_secure/White%20Papers/VBP_TimeToPrepareIsNow.pdf [accessed 2015-09-16] [WebCite Cache ID 6bZg5Noz3]
6. Porter ME. What is value in health care? N Engl J Med 2010 Dec 23;363(26):2477-2481. [doi: 10.1056/NEJMp1011024] [Medline: 21142528]
7. Deshpande PR, Rajan S, Sudeepthi BL, Abdul Nazir CP. Patient-reported outcomes: a new era in clinical research. Perspect Clin Res 2011 Oct;2(4):137-144 [FREE Full text] [doi: 10.4103/2229-3485.86879] [Medline: 22145124]
8. Manary MP, Boulding W, Staelin R, Glickman SW. The patient experience and health outcomes. N Engl J Med 2013 Jan 17;368(3):201-203. [doi: 10.1056/NEJMp1211775] [Medline: 23268647]
9. Jackson JL, Chamberlin J, Kroenke K. Predictors of patient satisfaction. Soc Sci Med 2001 Feb;52(4):609-620. [Medline: 11206657]
10. Hussey PS, Wertheimer S, Mehrotra A. The association between health care quality and cost: a systematic review. Ann Intern Med 2013 Jan 1;158(1):27-34. [doi: 10.7326/0003-4819-158-1-201301010-00006] [Medline: 23277898]
11. Fenton JJ, Jerant AF, Bertakis KD, Franks P. The cost of satisfaction: a national study of patient satisfaction, health care utilization, expenditures, and mortality. Arch Intern Med 2012 Mar 12;172(5):405-411. [doi: 10.1001/archinternmed.2011.1662] [Medline: 22331982]
12. Cohen JW, Monheit AC, Beauregard KM, Cohen SB, Lefkowitz DC, Potter DE, et al. The Medical Expenditure Panel Survey: a national health information resource. Inquiry 1996;33(4):373-389. [Medline: 9031653]
13. Sitzia J, Wood N. Patient satisfaction: a review of issues and concepts. Soc Sci Med 1997 Dec;45(12):1829-1843. [Medline: 9447632]
14. Hekkert KD, Cihangir S, Kleefstra SM, van den Berg B, Kool RB. Patient satisfaction revisited: a multilevel approach. Soc Sci Med 2009 Jul;69(1):68-75. [doi: 10.1016/j.socscimed.2009.04.016] [Medline: 19446942]

15. Sitzia J, Wood N. Patient satisfaction: a review of issues and concepts. Soc Sci Med 1997 Dec;45(12):1829-1843. [Medline: 9447632]
16. Stewart M, Brown JB, Donner A, McWhinney IR, Oates J, Weston WW, et al. The impact of patient-centered care on outcomes. J Fam Pract 2000 Sep;49(9):796-804. [Medline: 11032203]
17. Rahmqvist M, Bara A. Patient characteristics and quality dimensions related to patient satisfaction. Int J Qual Health Care 2010 Apr;22(2):86-92 [FREE Full text] [doi: 10.1093/intqhc/mzq009] [Medline: 20133477]
18. Hall JA, Dornan MC. Patient sociodemographic characteristics as predictors of satisfaction with medical care: a meta-analysis. Soc Sci Med 1990;30(7):811-818. [Medline: 2138357]
19. Young GJ, Meterko M, Desai KR. Patient satisfaction with hospital care: effects of demographic and institutional characteristics. Med Care 2000 Mar;38(3):325-334. [Medline: 10718357]
20. Medical Expenditure Panel Survey. Road Rockville, MD: Agency for Healthcare Research and Quality; 2011. National Health Care Expenses in the U.S. Civilian Noninstitutionalized Population URL: http://meps.ahrq.gov/mepsweb/data_stats/Pub_ProdResults_Details.jsp?pt=Statistical%20Brief&opt=2&id=1125 [WebCite Cache ID 6bcjfpUpI]
21. Covinsky KE, Rosenthal GE, Chren MM, Justice AC, Fortinsky RH, Palmer RM, et al. The relation between health status changes and patient satisfaction in older hospitalized medical patients. J Gen Intern Med 1998 Apr;13(4):223-229 [FREE Full text] [Medline: 9565384]
22. Manary MP, Boulding W, Staelin R, Glickman SW. The patient experience and health outcomes. N Engl J Med 2013 Jan 17;368(3):201-203. [doi: 10.1056/NEJMp1211775] [Medline: 23268647]
23. Bowling A, Rowe G, Lambert N, Waddington M, Mahtani K, Kenten C, et al. The measurement of patients' expectations for health care: a review and psychometric testing of a measure of patients' expectations. Health Technol Assess 2012 Jul;16(30):i-xii, 1 [FREE Full text] [doi: 10.3310/hta16300] [Medline: 22747798]
24. Fu AZ, Wang N. Healthcare expenditures and patient satisfaction: cost and quality from the consumer's perspective in the US. Curr Med Res Opin 2008 May;24(5):1385-1394. [doi: 10.1185/030079908X291994] [Medline: 18387219]

Abbreviations

CAHPS: Consumer Assessment of Healthcare Providers and Systems
CMS: Center for Medicare & Medicaid Services
EXP: total health care expenditure
GH: general health
MCS: mental health component score
MEPS: Medical Expenditures Panel Survey
PCS: physical health component score
PRO: patient-reported outcomes
SAT: patient satisfaction

Edited by G Eysenbach; submitted 19.02.15; peer-reviewed by E Papastavrou, M Rahmqvist; comments to author 07.04.15; revised version received 12.07.15; accepted 17.08.15; published 23.09.15.

Please cite as:
Hung M, Zhang W, Chen W, Bounsanga J, Cheng C, Franklin JD, Crum AB, Voss MW, Hon SD
Patient-Reported Outcomes and Total Health Care Expenditure in Prediction of Patient Satisfaction: Results From a National Study
JMIR Public Health Surveill 2015;1(2):e13
URL: http://publichealth.jmir.org/2015/2/e13/
doi:10.2196/publichealth.4360
PMID:27227131

Publisher:
JMIR Publications
130 Queens Quay East.
Toronto, ON, M5A 3Y5
Phone: (+1) 416-583-2040
Email: support@jmir.org