

---

# JMIR Public Health and Surveillance

---

Impact Factor (2022): 8.5  
Volume 2 (2016), Issue 2 ISSN 2369-2960 Editor in Chief: Travis Sanchez, PhD, MPH

---

## Contents

### Review

- The Use of Social Media for Health Promotion in Hispanic Populations: A Scoping Systematic Review (e32)  
Julia Hudnut-Beumler, Eli Po'e, Shari Barkin. . . . . 4

### Original Papers

- Testing the Feasibility of a Passive and Active Case Ascertainment System for Multiple Rare Conditions Simultaneously: The Experience in Three US States (e151)  
Amanda Reichard, Suzanne McDermott, Margaret Rutenber, Joshua Mann, Michael Smith, Julie Royer, Rodolfo Valdez. . . . . 15
- IBM Watson Analytics: Automating Visualization, Descriptive, and Predictive Statistics (e157)  
Robert Hoyt, Dallas Snider, Carla Thompson, Sarita Mantravadi. . . . . 27
- Efficacy of Web-Based Collection of Strength-Based Testimonials for Text Message Extension of Youth Suicide Prevention Program: Randomized Controlled Experiment (e164)  
Phyo Thiha, Anthony Pisani, Kunal Gurditta, Erin Cherry, Derick Peterson, Henry Kautz, Peter Wyman. . . . . 39
- Perceptions Toward a Smoking Cessation App Targeting LGBTQ+ Youth and Young Adults: A Qualitative Framework Analysis of Focus Groups (e165)  
N Baskerville, Darly Dash, Katy Wong, Alanna Shuh, Aneta Abramowicz. . . . . 50
- Reaction on Twitter to a Cluster of Perinatal Deaths: A Mixed Method Study (e36)  
Sarah Meaney, Leanne Cussen, Richard Greene, Keelin O'Donoghue. . . . . 61
- Chiropractic and Spinal Manipulation Therapy on Twitter: Case Study Examining the Presence of Critiques and Debates (e153)  
Alessandro Marcon, Philip Klostermann, Timothy Caulfield. . . . . 73
- Building a National Neighborhood Dataset From Geotagged Twitter Data for Indicators of Happiness, Diet, and Physical Activity (e158)  
Quynh Nguyen, Dapeng Li, Hsien-Wen Meng, Suraj Kath, Elaine Nsoesie, Feifei Li, Ming Wen. . . . . 82
- Evaluating Google, Twitter, and Wikipedia as Tools for Influenza Surveillance Using Bayesian Change Point Analysis: A Comparative Analysis (e161)  
J Sharpe, Richard Hopkins, Robert Cook, Catherine Striley. . . . . 98

<p><b>"When 'Bad' is 'Good'": Identifying Personal Communication and Sentiment in Drug-Related Tweets (e162)</b>          Raminta Daniulaityte, Lu Chen, Francois Lamy, Robert Carlson, Krishnaprasad Thirunarayan, Amit Sheth. . . . .</p>	111
<p><b>Estimating the Duration of Public Concern After the Fukushima Dai-ichi Nuclear Power Station Accident From the Occurrence of Radiation Exposure-Related Terms on Twitter: A Retrospective Data Analysis (e168)</b>          Naoki Nishimoto, Mizuki Ota, Ayako Yagahara, Katsuhiko Ogasawara. . . . .</p>	123
<p><b>E-Cigarette Social Media Messages: A Text Mining Analysis of Marketing and Consumer Conversations on Twitter (e171)</b>          Allison Lazard, Adam Saffer, Gary Wilcox, Arnold Chung, Michael Mackert, Jay Bernhardt. . . . .</p>	137
<p><b>Disease Monitoring and Health Campaign Evaluation Using Google Search Activities for HIV and AIDS, Stroke, Colorectal Cancer, and Marijuana Use in Canada: A Retrospective Observational Study (e156)</b>          Rebecca Ling, Joon Lee. . . . .</p>	146
<p><b>"Mommy Blogs" and the Vaccination Exemption Narrative: Results From A Machine-Learning Approach for Story Aggregation on Parenting Social Media Sites (e166)</b>          Timothy Tangherlini, Vwani Roychowdhury, Beth Glenn, Catherine Crespi, Roja Bandari, Akshay Wadia, Misagh Falahi, Ehsan Ebrahimzadeh, Roshan Bastani. . . . .</p>	161
<p><b>A Comparison of Web and Telephone Responses From a National HIV and AIDS Survey (e37)</b>          Marcella Jones, Liviana Calzavara, Dan Allman, Catherine Worthington, Mark Tyndall, James Iveniuk. . . . .</p>	176
<p><b>Acceptability of a Community-Based Outreach HIV-Testing Intervention Using Oral Fluid Collection Devices and Web-Based HIV Test Result Collection Among Sub-Saharan African Migrants: A Mixed-Method Study (e33)</b>          Jasna Loos, Lazare Manirankunda, Tom Platteau, Laura Albers, Katrien Fransen, Tine Vermoesen, Fiona Namanya, Christiana Nöstlinger. . . . .</p>	191
<p><b>Determining Survey Satisficing of Online Longitudinal Survey Data in the Multicenter AIDS Cohort Study: A Group-Based Trajectory Analysis (e150)</b>          Junrui Di, Ying Li, M Friedman, Susheel Reddy, Pamela Surkan, Steven Shoptaw, Michael Plankey. . . . .</p>	204
<p><b>How Best to Obtain Valid, Verifiable Data Online From Male Couples? Lessons Learned From an eHealth HIV Prevention Intervention for HIV-Negative Male Couples (e152)</b>          Jason Mitchell, Ji-Young Lee, Rob Stephenson. . . . .</p>	214
<p><b>Investigating Sociodemographic Factors and HIV Risk Behaviors Associated With Social Networking Among Adolescents in Soweto, South Africa: A Cross-Sectional Survey (e154)</b>          Janan Dietrich, Fatima Laher, Stefanie Hornschuh, Busisiwe Nkala, Lucy Chimoyi, Kennedy Otwombe, Angela Kaida, Glenda Gray, Cari Miller. . . . .</p>	222
<p><b>A Test of Concept Study of At-Home, Self-Administered HIV Testing With Web-Based Peer Counseling Via Video Chat for Men Who Have Sex With Men (e170)</b>          Jessica Maksut, Lisa Eaton, Elizabeth Siembida, Daniel Driffin, Robert Baldwin. . . . .</p>	235
<p><b>Exploring Consumer and Patient Knowledge, Behavior, and Attitude Toward Medicinal and Lifestyle Products Purchased From the Internet: A Web-Based Survey (e34)</b>          Sulaf Assi, Jordan Thomas, Mohamed Haffar, David Osselton. . . . .</p>	258
<p><b>Development of an Ecological Momentary Assessment Mobile App for a Low-Literacy, Mexican American Population to Collect Disordered Eating Behaviors (e31)</b>          Kay Connelly, Karen Stein, Beenish Chaudry, Nicole Trabold. . . . .</p>	301

**Qualitative Analysis of Cognitive Interviews With School Children: A Web-Based Food Intake Questionnaire (e167)**  
 Vanessa Fernandes Davies, Emil Kupek, Patricia Faria Di Pietro, Maria Altenburg de Assis, Francilene GK Vieira, Clarice Perucchi, Rafaella Mafra, Debbe Thompson, Thomas Baranowski. . . . . 315

**Effect of Performance Feedback on Community Health Workers' Motivation and Performance in Madhya Pradesh, India: A Randomized Controlled Trial (e169)**  
 Sangya Kaphle, Michael Matheke-Fischer, Neal Lesh. . . . . 330

**Viewpoint**

**Lessons Learned From Methodological Validation Research in E-Epidemiology (e160)**  
 Emmanuelle Kesse-Guyot, Karen Assmann, Valentina Andreeva, Katia Castetbon, Caroline Méjean, Mathilde Touvier, Benoît Salanave, Valérie Deschamps, Sandrine Péneau, Léopold Fezeu, Chantal Julia, Benjamin Allès, Pilar Galan, Serge Hercberg. . . . . 245

**Protocol**

**The Prevalence of Sexual Behavior Stigma Affecting Gay Men and Other Men Who Have Sex with Men Across Sub-Saharan Africa and in the United States (e35)**  
 Shauna Stahlman, Travis Sanchez, Patrick Sullivan, Sosthenes Ketende, Carrie Lyons, Manhattan Charurat, Fatou Drame, Daouda Diouf, Rebecca Ezouatchi, Seni Kouanda, Simplice Anato, Tampose Mothopeng, Zandile Mnisi, Stefan Baral. . . . . 274

**Short Paper**

**High Burden of Unrecognized Atrial Fibrillation in Rural India: An Innovative Community-Based Cross-Sectional Screening Program (e159)**  
 Apurv Soni, Allison Earon, Anna Handorf, Nisha Fahey, Kandarp Talati, John Bostrom, Ki Chon, Craig Napolitano, Michael Chin, John Sullivan, Shyamsundar Raithatha, Robert Goldberg, Somashekhar Nimbalkar, Jeroan Allison, Sunil Thanvi, David McManus. . . . . 293

**Corrigenda and Addendas**

**Metadata Correction of: How Best to Obtain Valid, Verifiable Data Online From Male Couples? Lessons Learned From an eHealth HIV Prevention Intervention for HIV-Negative Male Couples (e155)**  
 Jason Mitchell, Ji-Young Lee, Rob Stephenson. . . . . 326

**Metadata Correction of: Lessons Learned From Methodological Validation Research in E-Epidemiology (e160)**  
 Emmanuelle Kesse-Guyot, Karen Assmann, Valentina Andreeva, Katia Castetbon, Caroline Méjean, Mathilde Touvier, Benoît Salanave, Valérie Deschamps, Sandrine Péneau, Léopold Fezeu, Chantal Julia, Benjamin Allès, Pilar Galan, Serge Hercberg. . . . . 328

---

Review

# The Use of Social Media for Health Promotion in Hispanic Populations: A Scoping Systematic Review

---

Julia Hudnut-Beumler<sup>1</sup>, BA; Eli Po'e<sup>2</sup>, BS; Shari Barkin<sup>2</sup>, MD, MSHS

<sup>1</sup>Vanderbilt University School of Medicine, Nashville, TN, United States

<sup>2</sup>Department of Pediatrics, Vanderbilt University School of Medicine, Nashville, TN, United States

**Corresponding Author:**

Shari Barkin, MD, MSHS

Department of Pediatrics, Vanderbilt University School of Medicine

2200 Children's Way

Doctor's Office Tower 8232

Nashville, TN, 37232-9225

United States

Phone: 1 615 322 7080

Fax: 1 615 936 1730

Email: [shari.barkin@vanderbilt.edu](mailto:shari.barkin@vanderbilt.edu)

---

## Abstract

**Background:** The Internet is an increasingly popular platform for public health interventions due to its distinct ability to communicate with, engage, and educate communities. Given the widespread use of the Internet, these interventions could be a means of equalizing access to information to address health disparities in minority populations, such as Hispanics. Hispanics are disproportionately affected by poor health outcomes, including obesity, diabetes, and human immunodeficiency virus/acquired immune deficiency syndrome. Although underserved and underrepresented, Hispanics are among the leading users of social media in the United States. Previous reviews have examined the use of social media in public health efforts, but, to our knowledge, none have focused on the Hispanic population.

**Objective:** To conduct a scoping systematic review of the published literature to capture the ways social media has been used in health interventions aimed at Hispanic populations and identify gaps in existing knowledge to provide recommendations for future research.

**Methods:** We performed a systematic review of the literature related to social media, public health, and Hispanics using the PubMed, PsycINFO, and EMBASE databases to locate peer-reviewed studies published between January 1, 2010, and December 31, 2015. Each article was reviewed for the following inclusion criteria: social media as a main component of study methodology or content; public health topic; majority Hispanic/Latino study population; English or Spanish language; and original research study. Relevant data were extracted from articles meeting inclusion criteria including publication year, location, study design, social media platform, use of social media, target population, and public health topic.

**Results:** Of the 267 articles retrieved, a total of 27 unique articles met inclusion criteria. All were published in 2012 or later. The most common study design was a cross-sectional survey, which was featured in 10 of the 27 (37%) articles. All articles used social media for at least one of the following three purposes: recruiting study participants (14 of 27, 52%), promoting health education (12 of 27, 44%), and/or describing social media users (12 of 27, 44%). All but one article used multiple social media platforms, though Facebook was by far the most popular appearing in 24 of the 27 (89%). A diverse array of Hispanic populations was targeted, and health topics featured. Of these, the most highly represented were articles on sexual health directed toward Latino men who have sex with men (12 of 27, 44%). Healthy eating and active living received the second greatest focus (4 of 27, 15%).

**Conclusions:** Social media offers a potential accessible venue for health interventions aimed at Hispanics, a group at disproportionate risk for poor health outcomes. To date, most publications are descriptive in nature, with few indicating specific interventions and associated outcomes to improve health.

(*JMIR Public Health Surveill* 2016;2(2):e32) doi:[10.2196/publichealth.5579](https://doi.org/10.2196/publichealth.5579)

---



**KEYWORDS**

social media; social networking; Hispanic Americans; public health; health behavior

## Introduction

The Internet is becoming an increasingly popular platform for public health interventions [1]. Key components of public health services, as outlined by the Center for Disease Control and Prevention, include educating communities about health issues, linking individuals to personal health services, and conducting research to find new insights and innovative solutions to health problems [2]. Compared with traditional forms of communication, the Internet has distinct advantages when it comes to communicating with, engaging, and educating communities [3]. For many individuals, the Internet is already their primary source of health information [1], and thus, it represents an opportune method of communication for public health interventions. Not only do Internet interventions have a high potential reach, but they have other advantages as well. Internet-based implementation allows individuals to access the intervention at a time that is convenient for them. That intervention can be highly personalized, based on participant data, and engaging, based on interactive tools and graphically rich content [1]. Additional benefits of using the Internet include low cost and rapid transmission through a wide community [4].

Public health practitioners and researchers alike have taken note of the potential the Internet holds for health interventions. Given the widespread use of the Internet, these interventions could be a means of addressing health disparities. Gibbons et al wrote of “the potential to connect underserved and underrepresented populations to important health information resources and to build social support for those affected by health care issues” through the use of the Internet [5]. Among classically vulnerable populations, such as rural, low-income, and racial and ethnic minorities [6], Hispanics are notable in that they “own smartphones, go online from a mobile device, and use social networking sites (SNSs) at similar—and sometimes higher—rates than do other groups of Americans” [7]. With regard to Internet use, 81% of Hispanic adults in the United States were on the Web in 2015, which is similar to the figure for all American adults of 84% [8]. However, when it comes to social media sites, Hispanics are among the leading users—75% of Hispanic adult Internet users were on Facebook in 2015, compared with 70% of white and 67% of black adult Internet users [9]. Hispanics are the leading users of other social media platforms as well, including Twitter and Pinterest [9]. Given the affinity of the Hispanic community for social media sites, they would appear to hold great potential for health interventions in this population. Moreover, Hispanics are at disproportionate risk for poor health outcomes, such as obesity, diabetes, and human immunodeficiency virus (HIV)/acquired immune deficiency syndrome (AIDS) [10-12].

Social media generally refers to Internet sites that allow users to generate and share content, which reached mainstream prominence in the early 2000s [13,14]. Examples include blogs, wikis, SNSs (eg, Facebook and Twitter), and content communities (eg, YouTube) [2,15]. Most previously published systematic reviews of social media-driven public health

interventions focus on a particular health topic such as weight management, healthy lifestyle behaviors, or smoking cessation [16-18]. In 2012, Capurro et al performed a systematic review to capture the ways SNSs had been used for public health research and practice. Although their review included all health topics, none of the articles included specifically mentioned or focused on Hispanics [3]. To our knowledge, no other review has been conducted to assess the ways social media has been used for public health promotion and research among the Hispanic population in the United States, even though this is one of the fastest growing demographics in America [19,20]. Given the popularity of social media sites among the Hispanic community, the use of social media sites for public health efforts in other populations, and the disproportionate poor health outcomes in this population [3,9-12], the goals of this review were to inform public health practitioners and researchers on the current state of knowledge in the field and highlight gaps in need of further study and research. In this review, Hispanic and Latino/a have been used interchangeably.

## Methods

### Literature Search

We conducted a literature search using 3 databases: PubMed, PsycINFO, and EMBASE. Our search terms included “Hispanic,” “Latino,” and “Latina,” “social media,” “social network,” “social networking,” and “social network site,” as well as specific SNS names. Because SNSs are constantly evolving and changing in popularity, we used Wikipedia’s list of SNSs current to the first date of search—December 1, 2015. Of the 209 sites listed, 4 generated additional results within our query. The others were eliminated from the search terms. Using this process, our PubMed search included: (“Social Media”[Mesh] OR “social media”[tiab] OR “social network”[tiab] OR “social networking”[tiab] OR “social network site”[tiab] OR “facebook”[tiab] OR “google+”[tiab] OR “myspace”[tiab] OR “youtube”[tiab]) AND (“Hispanic Americans”[Mesh] OR “Hispanic”[tiab] OR “Hispanics”[tiab] OR “Latino”[tiab] OR “Latina”[tiab] OR “Latinos”[tiab] OR “Latinas”[tiab]) AND (“Public Health”[Mesh] OR “public health”[tiab] OR “Health Behavior”[Mesh] OR “health behavior”[tiab]).

For the other 2 databases, the query was modified to fit their search specifications. From the articles identified, we conducted a manual search of the articles’ references for other relevant studies.

### Article Selection

Inclusion criteria included were (1) social media as a main component of study methodology or content; (2) public health topic (ie, disease prevention or health promotion efforts); (3) majority Hispanic/Latino study population (ie,  $\geq 50\%$ ); (4) published between January 1, 2010, and December 31, 2015; and (5) original research study.

We restricted our search to articles published in the last 5 years given the rapidly changing nature and popularity of Internet-based technology. We excluded studies not published in English or Spanish and duplicate articles. Two reviewers independently assessed each article based on these inclusion and exclusion criteria.

### Data Extraction

After identifying studies that met inclusion criteria, the following data were extracted from each article by 2 independent reviewers: publication year, geographic location, study design, sample size, study purpose, social media type (eg, Facebook, Twitter, YouTube, and so forth), social media use (eg, recruiting study participants, promoting public health messaging, and so forth), target population (eg, age, gender, sexual orientation, and so forth), and public health topic. Data were recorded in 2 unique spreadsheets. Based on observed trends, qualitative data syntheses were performed. An inductive approach was taken to data abstraction in which specific observations were then broadened to general conclusions.

## Results

### Articles Retrieved

We conducted our literature search in December 2015. After our initial search on December 1, we performed weekly searches of all 3 databases to identify any newly published articles. Our final search was on December 31. Both of the reviewers used the same set of searches performed and saved during this period to conduct their independent reviews. A total of 267 articles were returned from the 3 databases: 149 from PubMed; 30 from PsycINFO; and 88 from EMBASE. Seventy-seven articles were found to be duplicates. For the remaining 190 unique articles, we reviewed the title and abstract for the selection criteria. If necessary, we also reviewed the full text of the article. Between the 2 reviewers, there was an agreement of 97% regarding which articles met inclusion criteria. Any discrepancies in group allocation or data extraction were resolved through systematic review and discussion to arrive at a consensus. In total, 26 articles met inclusion criteria. The references of these articles were then scanned, and one additional article identified was found to meet inclusion criteria. Thus, a total of 27 articles were included in the final review. See [Table 1](#) for a brief description of each of the articles included.

**Table 1.** Brief description of articles meeting inclusion criteria.

Author/year	Description	Social media site(s) used	Health topic
Graham 2012 [21]	Investigate efficacy of Web-based advertising for Spanish language smoking cessation website	MySpace Latino, MiGente, Website	Smoking cessation
Jaganath 2012 [22]	Describe the creation of a new, Facebook-based training curriculum for community leaders on HIV <sup>a</sup> prevention	Facebook, MySpace	Sexual health
Justice-Gardiner 2012 [23]	Compare a traditional media outreach campaign with a new media outreach campaign aimed at Hispanic cancer survivors and their families	Facebook, Twitter, Website	Cancer
Vyas 2012 [24]	Examine use of SMS <sup>b</sup> and social media for decreasing sexual risk taking among Latino youth	SMS, Facebook, YouTube, Twitter, MySpace	Sexual health
Young 2012[25]	Determine whether peer leaders can be recruited for a community-based health intervention using social media	Facebook, MySpace, Craigslist	Sexual health
Dixon-Gray 2013 [26]	Develop and implement a social marketing campaign to increase preconception health knowledge among second-generation Latinas	Facebook, MySpace, Website	Women's health
Landry 2013 [27]	Examine associations between new media use and sexual behaviors	SMS, unspecified social media	Sexual health
Tucker 2013 [28]	Examine the association between exposure to alcohol or other drug-related media and use of alcohol among adolescents	Facebook, MySpace	Substance abuse
Young 2013 [29]	Determine the feasibility and acceptability of using SNSs <sup>c</sup> to facilitate HIV-related discussions and requests for home-based HIV test kits in Latino MSMs <sup>d</sup>	Facebook, MySpace, Craigslist	Sexual health
Young 2013 [30]	Understand the relationship between social media sex-seeking and sexual risk behaviors in Latino MSMs	Facebook, MySpace, Twitter, Grindr, Adam4Adam, Manhunt	Sexual health
Young 2013 [31]	Determine the feasibility of recruiting peer leaders for a community-based health intervention using social media	Facebook, MySpace	Sexual health
Young 2013 [32]	Explore the feasibility of recruiting minority MSM Facebook users for HIV prevention studies	Facebook Craigslist, Website	Sexual health
Young 2013 [33]	Determine whether social networking communities can increase HIV testing in Latino MSMs	Facebook, Craigslist, Website	Sexual health
Young 2013 [34]	Explore associations between stimulant use, sexual risk behaviors, and social networking among Latino MSMs	Facebook, Craigslist	Substance abuse
Barrera 2014 [35]	Examine comparative impact of keywords in an Web-based campaign to recruit pregnant Latina women to an Internet intervention	Website	Women's health
Ferguson 2014 [36]	Examine television, social media, and peer competition influences on body dissatisfaction and eating disorder symptoms in Hispanic adolescent females	Facebook, Twitter, YouTube, Wordpress, multiplayer online gaming sites	Body image and eating disorders
Hanson 2014 [37]	Determine use of social media for health care-related purposes among medically underserved primary care patients	Facebook, MySpace, LinkedIn, YouTube, Twitter, blogs, SMS texting	Patient-provider communication
Martinez 2014 [38]	Recruit Spanish-speaking, Latino gay couples with social media to an HIV prevention study	Facebook, Craigslist, Grindr, SCRUFF, Jack'd, Instagram, SMS, Website	Sexual health
Quintiliani 2014 [39]	Improve weight, diet, and physical activity among low socioeconomic status public housing residents with social media campaign component to intervention	SMS, Facebook	Healthy eating and active living
Young 2014 [40]	Describe study retention among Latino MSMs 1 year after a 12-week, social networking-based HIV prevention trial	Facebook, MySpace, Craigslist	Sexual health
Young 2014 [41]	Assess the feasibility and acceptability of using SNSs as a health research platform among Latino MSMs	Facebook, MySpace, Craigslist	Healthy eating and active living
Chiu 2015 [42]	Assess association between HIV status, SNS use, and sexual risk behaviors among Latino MSMs	Facebook, MySpace, Grindr, Orkut	Sexual health

Author/year	Description	Social media site(s) used	Health topic
Criss 2015 [43]	Explore how health information sources inform decision-making among Hispanic mothers	Facebook, YouTube, Website	Women's health
Frerichs 2015 [44]	Assess readiness of community to address obesity and adopt healthy lifestyles following social media campaign	Facebook, YouTube, Website	Healthy eating and active living
Lee 2015 [45]	Examine the association between Web-based health information-seeking behaviors and health behaviors	Facebook, Twitter, MySpace, Website	Healthy eating and active living
Price 2015 [46]	Summarize public awareness activities on epilepsy to highlight communication channels	Facebook, Twitter, Website	Epilepsy
Smaldone 2015 [47]	Assess Latino parent and adolescent preferences in the use of mobile technologies and social media for provider-patient communication	Facebook, Twitter, MySpace, AIM, iChat, Skype, Oovoo, SMS	Patient-provider communication

<sup>a</sup>HIV, human immunodeficiency virus.

<sup>b</sup>SMS, short message service.

<sup>c</sup>SNS, social networking sites.

<sup>d</sup>MSM, men who have sex with men.

### Trends in Publication Dates and Study Locations

All articles meeting inclusion criteria were published in the last 4 years, from 2012 to 2015. In 2012, there were 5 articles published on the topic. That number increased to 9 in 2013, 7 in 2014, and 6 in 2015. No earlier articles retrieved in the search met inclusion criteria. With regard to study location, 26 of the 27 articles (96%) detailed studies that were conducted in the United States. The one remaining study was conducted internationally with most participants coming from Central and South America.

### Social Media Platforms

Of the social media platforms used, Facebook was by far the most common, appearing in 24 of the 27 articles (89%). Other popular platforms included MySpace (56%), websites (44%), Twitter (30%), Craigslist (26%), and YouTube (19%). Fourteen other social media platforms, such as Instagram, Grindr, MiGente, and LinkedIn, were mentioned in the articles, although with lesser frequency. All but one study used multiple social media platforms.

### Types of Studies

The most common study design was a cross-sectional survey, which was featured in 10 of the 27 articles (37%). Most surveys were used to capture social media usage, users' characteristics, or users' communication preferences. The next most common study designs were qualitative studies and randomized control

trials. Qualitative studies appeared in 4 of the 27 articles (15%), and randomized control trials were featured in an equal number (4 of 27, 15%). Of note, 2 of the 4 papers featuring randomized control trials are from the same study, the Harnessing Online Peer Education study, meaning that in total there were only 2 unique randomized control trials identified in this review. Of the remaining studies, there were 3 (11%) prospective cohort studies, 3 (11%) mixed methods studies, and 3 (11%) reports on public health campaigns.

### Uses of Social Media

All of the articles included in this review used social media for at least one of the following 3 purposes: recruiting study participants, promoting health education, or describing users/usage characteristics. Most articles (14 of 27, 52%) used social media to identify and recruit study participants. However, using social media for health education (12 of 27, 44%) and describing the characteristics of social media users were both also highly represented in the articles (12 of 27, 44%). Though less common, a few studies assessed participants' health communication preferences (3 of 27, 11%), specifically whether they were open to receiving health information, including general public health messages and personal messages from health care providers, via social media. Finally, in 2 of the articles (7%), social media was highlighted as a way to retain study participants. See [Table 2](#) for uses of social media by target population and health topic.

**Table 2.** Uses of social media for health interventions by targeted Hispanic population and health topic.

	Uses of social media <sup>a</sup>				
	Recruiting study participants (n=14)	Promoting health education (n=12)	Describing users or usage characteristics (n=12)	Assessing communication preferences (n=3)	Retaining study participants (n=2)
Target Hispanic population <sup>b</sup>					
Men who have sex with men (n=12)	12 (86%)	7 (58%)	4 (33%)	0 (0%)	1 (50%)
Adolescents (n=5)	0 (0%)	0 (0%)	4 (33%)	2 (67%)	0 (0%)
General population (n=3)	0 (0%)	2 (17%)	1 (8%)	0 (0%)	0 (0%)
Pregnant women and young mothers (n=3)	1 (7%)	1 (8%)	1 (8%)	0 (0%)	0 (0%)
Medically underserved (n=1)	0 (0%)	0 (0%)	2 (17%)	1 (33%)	0 (0%)
Public housing residents (n=1)	0 (0%)	1 (8%)	0 (0%)	0 (0%)	1 (50%)
Cancer survivors (n=1)	0 (0%)	1 (8%)	0 (0%)	0 (0%)	0 (0%)
Smokers (n=1)	1 (7%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)
Health topic <sup>c</sup>					
Sexual health (n=12)	10 (71%)	6 (50%)	5 (42%)	1 (33%)	1 (50%)
Healthy eating and active living (n=4)	1 (7%)	3 (25%)	1 (8%)	0 (0%)	1 (50%)
Women's health (n=3)	1 (7%)	1 (8%)	1 (8%)	0 (0%)	0 (0%)
Substance abuse (n=3)	1 (7%)	0 (0%)	2 (17%)	0 (0%)	0 (0%)
Patient-provider communication (n=2)	0 (0%)	0 (0%)	2 (17%)	2 (67%)	0 (0%)
Body image and eating disorders (n=1)	0 (0%)	0 (0%)	1 (8%)	0 (0%)	0 (0%)
Cancer (n=1)	0 (0%)	1 (8%)	0 (0%)	0 (0%)	0 (0%)
Epilepsy (n=1)	0 (0%)	1 (8%)	0 (0%)	0 (0%)	0 (0%)
Smoking cessation (n=1)	1 (7%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)

<sup>a</sup>Articles may contain elements of multiple categories (n>27).

<sup>b</sup>Articles contain only one target Hispanic population (n=27).

<sup>c</sup>Articles contain only one health topic (n=27).

## Target Populations

With regard to target population, most articles focused on certain populations within the Hispanic community as opposed to individuals with specific medical conditions. An overwhelming number of the articles included in this review focused on Latino men who have sex with men (MSM). Forty-four percent (12 of 27) of articles targeted this population. However, 10 of these articles derived from a single study, the Harnessing Online Peer Education study. Other populations featured included adolescents (5 of 27, 19%), the medically underserved (1 of 27, 4%), public housing residents (1 of 27, 4%), and the general Hispanic population (3 of 27, 11%). Of these groups, the most commonly targeted were Hispanic adolescents and the general Hispanic population. Although less common, individuals with specific medical conditions were the focus of a few of the studies. These conditions included pregnant women and young mothers (3 of 27, 11%), cancer survivors (1 of 27, 4%), and smokers (1 of 27, 4%). Moreover, of the 14 articles that used social media to recruit study participants, 12 (86%) targeted MSM, followed distantly by pregnant women and young mothers (1 of 14, 7%) and smokers (1 of 14, 7%). None of the articles targeted at other populations used social media for

recruitment. Of the various groups targeted, MSM were also the most highly represented with regard to articles promoting health education (7 of 12, 58%), describing user characteristics (4 of 12, 33%), and retaining study participants (1 of 2, 50%).

## Health Topics

Of the health topics featured in the included articles, sexual health was by far the most highly represented, appearing in 12 of the 27 articles (44%). Two of these articles related to sexual health and risk taking behaviors among Hispanic adolescents. The remaining 10 articles were focused on sexual risk behaviors and HIV prevention among Latino MSMs. Other health topics that featured prominently were healthy eating and active living (4 of 27, 15%), women's health (3 of 27, 11%), and substance abuse (2 of 27, 7%). Additional topics discussed include patient-provider communication (2 of 27, 7%), body image and eating disorders (1 of 27; 4%), cancer (1 of 27; 4%), epilepsy (1 of 27; 4%), and smoking cessation (1 of 27; 4%). Of the health topics, sexual health featured prominently in articles using social media to recruit study participants (10 of 14, 71%), promote health education (6 of 12, 50%), describe user characteristics (5 of 12, 42%), and retain study participants (1 of 2, 50%).



## Discussion

### Principal Findings

Given Hispanic's disproportionate risk for poor health outcomes and high affinity for social media, public health interventions that use social media may be a means of addressing the health disparities present in this community [5,9-12]. With no articles published before 2012 and 27 articles published in the last 4 years relating to the use of social media for health interventions in Hispanic populations, this is evidently a nascent field in the realm of public health for this population. Of the 27 articles, 24 related to research, whereas only 3 were reports on public health campaigns, suggesting that using social media for health interventions in Hispanics is still mostly in the descriptive phase of research. It is not yet a common part of public health practice or published research. Although almost all articles referenced studies that included Facebook (89%), most used multiple social media platforms or complemented social media with other relatively new forms of communication, such as SMS or email. This trend deviates from previous reviews in which most studies depended on a single social media platform [3,17]. While text messaging on its own has received support as a tool for behavior change in disease prevention and management, the effectiveness of Internet interventions has been enhanced when coupled with additional methods of communication, especially SMS [48-50]. These distinct platforms may engage different aspects of the behavior change process, resulting in increased effectiveness of the intervention as a whole [50]. The use of multiple communication platforms by most studies included in this review may be a product of these findings, which come from previous trials that have sought to use the Internet to address health problems in other populations.

Perhaps given the relatively young nature of the field, few articles included in this review used randomized control trials. Four of the 27 articles featured randomized control trials; however, 2 of these articles come from the same study, suggesting that the lack of rigorous trial data is even less than it initially appears. Randomized control trials may also be underrepresented because of an inherent or perceived difficulty of using this study design for social media-based trials. Barrera et al noted the challenge in striking the right balance between addressing participant expectations of a trial and developing a social media-based study recruitment tool that captured the scope of the intervention without being too burdensome for Web users to access [35]. Moreover, few social media sites offer the option for closed-access or private groups. On most sites, content is open and freely available to everyone, which does not allow for the creation of intervention and control groups as would be necessary in a randomized control trial. Facebook is one exception. Part of Facebook's popularity in the articles included here may stem from the ability to create multiple private groups using the site, thus allowing for different exposure conditions.

Cross-sectional surveys and qualitative studies were the most highly represented study designs. These study designs align well with the way social media is currently being used for health interventions in the Hispanic population. While some studies

incorporated social media in multiple ways, every article included in this review used social media for at least one of the following 3 purposes: (1) study recruitment; , (2) health education promotion; , or (3) description of users/usage characteristics. Cross-sectional surveys and qualitative studies are appropriate for such descriptive purposes. Given the current gaps in knowledge, future studies should be designed as randomized control trials to test the effectiveness of social media-based health education programs and interventions for behavior change in Hispanic populations. Of the studies that have already been conducted, there is evidence that social media can be effective in recruiting study participants [25,31,32,38], increasing health knowledge [26], encouraging participants to engage in health discussions [33], increasing rates of testing for sexually transmitted diseases [29,33], and promoting study retention [40]. Only 2 studies specifically cited the use of social media as a means of promoting retention among study participants. Because of the challenges in retaining human subjects in longitudinal studies, this area merits continued study.

With regard to target study populations, a preponderance of articles focus on Latino MSMs. Because many of these articles come out of a single study, the number of studies that use social media-driven health interventions in Latino MSMs is actually far fewer. Three unique studies feature Latino MSMs—a similar number to that for Hispanic adolescents, Latinas who are pregnant or young mothers, and the Hispanic population at large. With only a handful of studies focused on each of these populations and fewer still on other groups like the medically underserved, continued research is needed for all Hispanic populations to further explore the efficacy and potential of social media-driven health interventions. Particular attention should be directed at the medically underserved and economically disadvantaged as a potential way to address health disparities among Hispanics [5]. Due to the large number of articles focusing on social media-based HIV prevention tools in Latino MSMs, sexual health is highly represented among the health topics featured in this review. However, once again, the actual number of unique studies dealing with sexual health is far fewer such that roughly an equivalent number of studies have been conducted to examine the use of social media in sexual health, healthy eating and active living, and women's health interventions. With only a few studies conducted on each topic, each of these merits further research, especially given the increased risk of poor health outcomes for Hispanic Americans compared with non-Hispanic, white Americans in these areas, specifically heightened risks for HIV/AIDS, obesity, diabetes, and late entry into or complete lack of prenatal care [10-12,51].

From the research that has been done, social media has been shown to be an accessible and effective method of engaging a variety of Hispanic populations in diverse health behavior change topics. Despite previous reports showing that Hispanics who are younger, more affluent, English-speaking, and native born are more likely to use social media [7-9], there is early evidence to suggest that social media can also be an effective platform for reaching other members of the Hispanic community who tend to fall outside of these categories, such as community health center patients [37]. For public health practitioners, social media holds great potential. There are already examples

community-specific efforts, such as Saludable Omaha which used Facebook and YouTube to promote healthy eating and active living among Omaha's Hispanic population [44], and nationwide campaigns, like that undertaken by the Epilepsy Foundation to increase public understanding and awareness of epilepsy among Hispanics through Facebook and Twitter [46]. Public health practitioners should both encourage the use of relevant pre-existing social media resources and develop new, social media-based public health campaigns for Hispanic populations.

### Limitations

A number of limitations do exist to this review. We limited our search to published, peer-reviewed journals, which could reflect a publication bias and does not include studies in progress but not yet published. In addition, we limited our search to articles published in the last 5 years, meaning that we may have missed

articles published earlier. Given that all articles meeting inclusion criteria were published after 2012, this is unlikely. Moreover, because of the rapidly changing nature and popularity of Internet-based technology, what may have been relevant and effective over 5 years ago may be outdated today.

### Conclusions

Hispanics are at increased risk for poor health outcomes, such as obesity, diabetes, and HIV/AIDS. Given that Hispanics are among the leading users of social media in the United States, there is an opportunity to use this technology for health interventions. Although a relatively new field, use of social media is quickly growing and a diverse array of target populations and health topics have been studied. To date, most publications are descriptive in nature and the full potential of social media-driven interventions for affecting health behavior change has yet to be fully realized in the Hispanic population.

### Acknowledgments

5 P30 DK92986-05 (PI: Elasy) supported Dr. Barkin's work on this manuscript. A portion of this project was supported by the Office of Medical Student Research at Vanderbilt University's School of Medicine.

### Conflicts of Interest

None declared.

### References

- Bennett GG, Glasgow RE. The delivery of public health interventions via the Internet: actualizing their potential. *Annu Rev Public Health* 2009 Apr;30:273-292. [doi: [10.1146/annurev.publhealth.031308.100235](https://doi.org/10.1146/annurev.publhealth.031308.100235)] [Medline: [19296777](https://pubmed.ncbi.nlm.nih.gov/19296777/)]
- Centers for Disease Control and Prevention. 2014 May 29. The Public Health System and the 10 Essential Public Health Services URL: <http://www.cdc.gov/nphsp/essentialservices.html> [accessed 2016-01-17] [WebCite Cache ID 6ecB3Bj9t]
- Capurro D, Cole K, Echavarría MI, Joe J, Neogi T, Turner AM. The use of social networking sites for public health practice and research: a systematic review. *J Med Internet Res* 2014 Mar;16(3):e79 [FREE Full text] [doi: [10.2196/jmir.2679](https://doi.org/10.2196/jmir.2679)] [Medline: [24642014](https://pubmed.ncbi.nlm.nih.gov/24642014/)]
- Vance K, Howe W, Dellavalle RP. Social internet sites as a source of public health information. *Dermatol Clin* 2009 Apr;27(2):133-6, vi. [doi: [10.1016/j.det.2008.11.010](https://doi.org/10.1016/j.det.2008.11.010)] [Medline: [19254656](https://pubmed.ncbi.nlm.nih.gov/19254656/)]
- Gibbons MC, Fleisher L, Slamon RE, Bass S, Kandadai V, Beck JR. Exploring the potential of Web 2.0 to address health disparities. *J Health Commun* 2011 Jul;16 Suppl 1:77-89. [doi: [10.1080/10810730.2011.596916](https://doi.org/10.1080/10810730.2011.596916)] [Medline: [21843097](https://pubmed.ncbi.nlm.nih.gov/21843097/)]
- Centers for Public Health Preparedness Network. 2007 Oct. 2006-2007 ASPH/CDC Vulnerable Populations Collaboration Group Preparedness Resource Kit URL: <http://preparedness.asph.org/perlc/documents/VulnerablePopulations.pdf> [accessed 2016-01-17] [WebCite Cache ID 6G2vrxcg2]
- Lopez MH, Gonzalez-Barrera A, Patten E. Pew Research Center. 2013 Mar 07. Closing the Digital Divide: Latinos and Technology Adoption URL: [http://www.pewhispanic.org/files/2013/03/Latinos\\_Social\\_Media\\_and\\_Mobile\\_Tech\\_03-2013\\_final.pdf](http://www.pewhispanic.org/files/2013/03/Latinos_Social_Media_and_Mobile_Tech_03-2013_final.pdf) [accessed 2016-01-17] [WebCite Cache ID 6ecA1rwuu]
- Perrin A, Duggan M. Pew Research Center. 2015 Jun 26. Americans' Internet Access: 2000-2015 URL: [http://www.pewinternet.org/files/2015/06/2015-06-26\\_internet-usage-across-demographics-discover\\_FINAL.pdf](http://www.pewinternet.org/files/2015/06/2015-06-26_internet-usage-across-demographics-discover_FINAL.pdf) [accessed 2016-01-17] [WebCite Cache ID 6ec4k7doj]
- Duggan M. Pew Research Center. 2015 Aug 19. Mobile Messaging and Social Media 2015 URL: <http://www.pewinternet.org/2015/08/19/mobile-messaging-and-social-media-2015/> [accessed 2016-01-17] [WebCite Cache ID 6ec3bGgMo]
- Schiller JS, Lucas JW, Peregoy JA. National Center for Health Statistics. 2012 Dec. Summary Health Statistics for U.S. Adults: National Health Interview Survey, 2011 URL: [http://www.cdc.gov/nchs/data/series/sr\\_10/sr10\\_256.pdf](http://www.cdc.gov/nchs/data/series/sr_10/sr10_256.pdf) [accessed 2016-01-26] [WebCite Cache ID 6eqvnSaJz]
- Blackwell DL, Lucas JW, Clarke TC. National Center for Health Statistics. 2014 Feb. Summary Health Statistics for U.S. adults: National Health Interview Survey, 2012 URL: [http://www.cdc.gov/nchs/data/series/sr\\_10/sr10\\_260.pdf](http://www.cdc.gov/nchs/data/series/sr_10/sr10_260.pdf) [accessed 2016-01-26] [WebCite Cache ID 6eqvzpJyT]
- Center for Disease Control and Prevention. 2013 Feb. HIV Surveillance Report, 2011 URL: [http://www.cdc.gov/hiv/surveillance/resources/reports/2011report/pdf/2011\\_HIV\\_Surveillance\\_Report\\_vol\\_23.pdf](http://www.cdc.gov/hiv/surveillance/resources/reports/2011report/pdf/2011_HIV_Surveillance_Report_vol_23.pdf) [accessed 2016-01-27] [WebCite Cache ID 6eqwIHdJ]



13. Kaplan AM, Haenlein M. Users of the world, unite! The challenges and opportunities of Social Media. *Business Horizons* 2010;53(1):59-68. [doi: [10.1016/j.bushor.2009.09.003](https://doi.org/10.1016/j.bushor.2009.09.003)]
14. Boyd DM, Ellison NB. Social Network Sites: Definition, History, and Scholarship. *J Comput Mediat Commun* 2008;13:210-230. [doi: [10.1111/j.1083-6101.2007.00393.x](https://doi.org/10.1111/j.1083-6101.2007.00393.x)]
15. Perrin A. Pew Research Center. 2015 Oct 08. Social Media Usage: 2005-2015 URL: [http://www.pewinternet.org/files/2015/10/PI\\_2015-10-08\\_Social-Networking-Usage-2005-2015\\_FINAL.pdf](http://www.pewinternet.org/files/2015/10/PI_2015-10-08_Social-Networking-Usage-2005-2015_FINAL.pdf) [accessed 2016-01-17] [WebCite Cache ID [6ec1hbiIQ](https://www.webcitation.org/6ec1hbiIQ)]
16. Bardus M, Smith JR, Samaha L, Abraham C. Mobile Phone and Web 2.0 Technologies for Weight Management: A Systematic Scoping Review. *J Med Internet Res* 2015;17(11):e259 [FREE Full text] [doi: [10.2196/jmir.5129](https://doi.org/10.2196/jmir.5129)] [Medline: [26573984](https://pubmed.ncbi.nlm.nih.gov/26573984/)]
17. Williams G, Hamm MP, Shulhan J, Vandermeer B, Hartling L. Social media interventions for diet and exercise behaviours: a systematic review and meta-analysis of randomised controlled trials. *BMJ Open* 2014 Feb;4(2):e003926 [FREE Full text] [doi: [10.1136/bmjopen-2013-003926](https://doi.org/10.1136/bmjopen-2013-003926)] [Medline: [24525388](https://pubmed.ncbi.nlm.nih.gov/24525388/)]
18. Jacobs MA, Cobb CO, Abrams L, Graham AL. Facebook apps for smoking cessation: a review of content and adherence to evidence-based guidelines. *J Med Internet Res* 2014 Sep;16(9):e205 [FREE Full text] [doi: [10.2196/jmir.3491](https://doi.org/10.2196/jmir.3491)] [Medline: [25205129](https://pubmed.ncbi.nlm.nih.gov/25205129/)]
19. Brown A. Pew Research Center. 2014 Jun 26. U.S. Hispanic and Asian populations growing, but for different reasons URL: <http://www.pewresearch.org/fact-tank/2014/06/26/u-s-hispanic-and-asian-populations-growing-but-for-different-reasons/> [accessed 2016-01-27] [WebCite Cache ID [6eqyIA2Be](https://www.webcitation.org/6eqyIA2Be)]
20. Colby SL, Ortman JM. United States Census Bureau. 2015 Mar. Projections of the Size and Composition of the U.S. Population: 2014 to 2060 URL: <https://www.census.gov/content/dam/Census/library/publications/2015/demo/p25-1143.pdf> [accessed 2016-01-27] [WebCite Cache ID [6eqyaCWeo](https://www.webcitation.org/6eqyaCWeo)]
21. Graham AL, Fang Y, Moreno JL, Streiff SL, Villegas J, Muñoz RF, et al. Online advertising to reach and recruit Latino smokers to an internet cessation program: impact and costs. *J Med Internet Res* 2012 Aug;14(4):e116 [FREE Full text] [doi: [10.2196/jmir.2162](https://doi.org/10.2196/jmir.2162)] [Medline: [22954502](https://pubmed.ncbi.nlm.nih.gov/22954502/)]
22. Jaganath D, Gill HK, Cohen AC, Young SD. Harnessing Online Peer Education (HOPE): integrating C-POL and social media to train peer leaders in HIV prevention. *AIDS Care* 2012 May;24(5):593-600 [FREE Full text] [doi: [10.1080/09540121.2011.630355](https://doi.org/10.1080/09540121.2011.630355)] [Medline: [22149081](https://pubmed.ncbi.nlm.nih.gov/22149081/)]
23. Justice-Gardiner H, Nutt S, Reche R, McMillan B, Warf R. Using new media to reach Hispanic/Latino cancer survivors. *J Cancer Educ* 2012 Mar;27(1):100-104. [doi: [10.1007/s13187-011-0267-2](https://doi.org/10.1007/s13187-011-0267-2)] [Medline: [21877170](https://pubmed.ncbi.nlm.nih.gov/21877170/)]
24. Vyas AN, Landry M, Schnider M, Rojas AM, Wood SF. Public health interventions: reaching Latino adolescents via short message service and social media. *J Med Internet Res* 2012 Jul;14(4):e99 [FREE Full text] [doi: [10.2196/jmir.2178](https://doi.org/10.2196/jmir.2178)] [Medline: [22789678](https://pubmed.ncbi.nlm.nih.gov/22789678/)]
25. Young SD. Analysis of online social networking peer health educators. *Stud Health Technol Inform* 2012;181:253-259. [Medline: [22954866](https://pubmed.ncbi.nlm.nih.gov/22954866/)]
26. Dixon-Gray LA, Mobley A, McFarlane JM, Rosenberg KD. Amor y Salud (Love and Health): a preconception health campaign for second-generation Latinas in Oregon. *Am J Health Promot* 2013 Jan;27(3 Suppl):S74-S76. [doi: [10.4278/ajhp.120113-ARB-29](https://doi.org/10.4278/ajhp.120113-ARB-29)] [Medline: [23286668](https://pubmed.ncbi.nlm.nih.gov/23286668/)]
27. Landry M, Gonzales FA, Wood S, Vyas A. New media use and sexual behavior among Latino adolescents. *Am J Health Behav* 2013 May;37(3):422-430. [doi: [10.5993/AJHB.37.3.15](https://doi.org/10.5993/AJHB.37.3.15)] [Medline: [23985189](https://pubmed.ncbi.nlm.nih.gov/23985189/)]
28. Tucker JS, Miles JN, D'Amico EJ. Cross-lagged associations between substance use-related media exposure and alcohol use during middle school. *J Adolesc Health* 2013 Oct;53(4):460-464 [FREE Full text] [doi: [10.1016/j.jadohealth.2013.05.005](https://doi.org/10.1016/j.jadohealth.2013.05.005)] [Medline: [23770074](https://pubmed.ncbi.nlm.nih.gov/23770074/)]
29. Young SD, Jaganath D. Online social networking for HIV education and prevention: a mixed-methods analysis. *Sex Transm Dis* 2013 Feb;40(2):162-167 [FREE Full text] [doi: [10.1097/OLQ.0b013e318278bd12](https://doi.org/10.1097/OLQ.0b013e318278bd12)] [Medline: [23324979](https://pubmed.ncbi.nlm.nih.gov/23324979/)]
30. Young SD, Szekeres G, Coates T. The relationship between online social networking and sexual risk behaviors among men who have sex with men (MSM). *PLoS One* 2013 May;8(5):e62271 [FREE Full text] [doi: [10.1371/journal.pone.0062271](https://doi.org/10.1371/journal.pone.0062271)] [Medline: [23658716](https://pubmed.ncbi.nlm.nih.gov/23658716/)]
31. Young SD, Harrell L, Jaganath D, Cohen AC, Shoptaw S. Feasibility of recruiting peer educators for an online social networking-based health intervention. *Health Educ J* 2013 May;72(3):276-282 [FREE Full text] [doi: [10.1177/0017896912440768](https://doi.org/10.1177/0017896912440768)] [Medline: [25530624](https://pubmed.ncbi.nlm.nih.gov/25530624/)]
32. Young SD, Szekeres G, Coates T. Sexual risk and HIV prevention behaviours among African-American and Latino MSM social networking users. *Int J STD AIDS* 2013 Aug;24(8):643-649 [FREE Full text] [doi: [10.1177/0956462413478875](https://doi.org/10.1177/0956462413478875)] [Medline: [23970575](https://pubmed.ncbi.nlm.nih.gov/23970575/)]
33. Young SD, Cumberland WG, Lee SJ, Jaganath D, Szekeres G, Coates T. Social networking technologies as an emerging tool for HIV prevention: a cluster randomized trial. *Ann Intern Med* 2013 Sep 3;159(5):318-324 [FREE Full text] [doi: [10.7326/0003-4819-159-5-201309030-00005](https://doi.org/10.7326/0003-4819-159-5-201309030-00005)] [Medline: [24026317](https://pubmed.ncbi.nlm.nih.gov/24026317/)]
34. Young SD, Shoptaw S. Stimulant use among African American and Latino MSM social networking users. *J Addict Dis* 2013 Jan;32(1):39-45 [FREE Full text] [doi: [10.1080/10550887.2012.759859](https://doi.org/10.1080/10550887.2012.759859)] [Medline: [23480246](https://pubmed.ncbi.nlm.nih.gov/23480246/)]

35. Barrera AZ, Kelman AR, Muñoz RF. Keywords to recruit Spanish- and English-speaking participants: evidence from an online postpartum depression randomized controlled trial. *J Med Internet Res* 2014 Jan;16(1):e6 [FREE Full text] [doi: [10.2196/jmir.2999](https://doi.org/10.2196/jmir.2999)] [Medline: [24407163](https://pubmed.ncbi.nlm.nih.gov/24407163/)]
36. Ferguson CJ, Muñoz ME, Garza A, Galindo M. Concurrent and prospective analyses of peer, television and social media influences on body dissatisfaction, eating disorder symptoms and life satisfaction in adolescent girls. *J Youth Adolesc* 2014 Jan;43(1):1-14. [doi: [10.1007/s10964-012-9898-9](https://doi.org/10.1007/s10964-012-9898-9)] [Medline: [23344652](https://pubmed.ncbi.nlm.nih.gov/23344652/)]
37. Hanson CL, West J, Thackeray R, Barnes MD, Downey J. Understanding and predicting social media use among community health center patients: a cross-sectional survey. *J Med Internet Res* 2014 Nov;16(11):e270 [FREE Full text] [doi: [10.2196/jmir.3373](https://doi.org/10.2196/jmir.3373)] [Medline: [25427823](https://pubmed.ncbi.nlm.nih.gov/25427823/)]
38. Martinez O, Wu E, Shultz AZ, Capote J, López RJ, Sandfort T, et al. Still a hard-to-reach population? Using social media to recruit Latino gay couples for an HIV intervention adaptation study. *J Med Internet Res* 2014 Apr;16(4):e113 [FREE Full text] [doi: [10.2196/jmir.3311](https://doi.org/10.2196/jmir.3311)] [Medline: [24763130](https://pubmed.ncbi.nlm.nih.gov/24763130/)]
39. Quintiliani LM, DeBiase MA, Branco JM, Bhosrekar SG, Rorie JA, Bowen DJ. Enhancing physical and social environments to reduce obesity among public housing residents: rationale, trial design, and baseline data for the Healthy Families study. *Contemp Clin Trials* 2014 Nov;39(2):201-210. [doi: [10.1016/j.cct.2014.08.005](https://doi.org/10.1016/j.cct.2014.08.005)] [Medline: [25139728](https://pubmed.ncbi.nlm.nih.gov/25139728/)]
40. Young SD. Social media technologies for HIV prevention study retention among minority men who have sex with men (MSM). *AIDS Behav* 2014 Sep;18(9):1625-1629 [FREE Full text] [doi: [10.1007/s10461-013-0604-z](https://doi.org/10.1007/s10461-013-0604-z)] [Medline: [24062015](https://pubmed.ncbi.nlm.nih.gov/24062015/)]
41. Young SD, Jaganath D. Feasibility of using social networking technologies for health research among men who have sex with men: a mixed methods study. *Am J Mens Health* 2014 Jan;8(1):6-14 [FREE Full text] [doi: [10.1177/1557988313476878](https://doi.org/10.1177/1557988313476878)] [Medline: [23407600](https://pubmed.ncbi.nlm.nih.gov/23407600/)]
42. Chiu CJ, Young SD. The Relationship Between Online Social Network Use, Sexual Risk Behaviors, and HIV Sero-Status Among a Sample of Predominately African American and Latino Men Who have Sex with Men (MSM) Social Media Users. *AIDS Behav* 2015 Jun;19 Suppl 2:98-105. [doi: [10.1007/s10461-014-0986-6](https://doi.org/10.1007/s10461-014-0986-6)] [Medline: [25572831](https://pubmed.ncbi.nlm.nih.gov/25572831/)]
43. Criss S, Woo Baidal JA, Goldman RE, Perkins M, Cunningham C, Taveras EM. The Role of Health Information Sources in Decision-Making Among Hispanic Mothers During Their Children's First 1000 Days of Life. *Matern Child Health J* 2015 Nov;19(11):2536-2543. [doi: [10.1007/s10995-015-1774-2](https://doi.org/10.1007/s10995-015-1774-2)] [Medline: [26122256](https://pubmed.ncbi.nlm.nih.gov/26122256/)]
44. Frerichs L, Brittin J, Robbins R, Steenson S, Stewart C, Fisher C, et al. SaludABLEOmaha: improving readiness to address obesity through healthy lifestyle in a Midwestern Latino community, 2011-2013. *Prev Chronic Dis* 2015 Feb;12:E20 [FREE Full text] [doi: [10.5888/pcd12.140328](https://doi.org/10.5888/pcd12.140328)] [Medline: [25674679](https://pubmed.ncbi.nlm.nih.gov/25674679/)]
45. Lee YJ, Boden-Albala B, Jia H, Wilcox A, Bakken S. The Association Between Online Health Information-Seeking Behaviors and Health Behaviors Among Hispanics in New York City: A Community-Based Cross-Sectional Study. *J Med Internet Res* 2015 Nov;17(11):e261 [FREE Full text] [doi: [10.2196/jmir.4368](https://doi.org/10.2196/jmir.4368)] [Medline: [26611438](https://pubmed.ncbi.nlm.nih.gov/26611438/)]
46. Price P, Kobau R, Buelow J, Austin J, Lowenberg K. Improving understanding, promoting social inclusion, and fostering empowerment related to epilepsy: Epilepsy Foundation public awareness campaigns--2001 through 2013. *Epilepsy Behav* 2015 Mar;44:239-244. [doi: [10.1016/j.yebeh.2014.12.044](https://doi.org/10.1016/j.yebeh.2014.12.044)] [Medline: [25726152](https://pubmed.ncbi.nlm.nih.gov/25726152/)]
47. Smaldone A, Stockwell MS, Osborne JC, Cortes Y, Bekele E, Green NS. Adolescent and parent use of new technologies for health communication: a study in an urban latino community. *J Public Health Res* 2015 Feb 19;4(1):376 [FREE Full text] [doi: [10.4081/jphr.2015.376](https://doi.org/10.4081/jphr.2015.376)] [Medline: [25918691](https://pubmed.ncbi.nlm.nih.gov/25918691/)]
48. Cole-Lewis H, Kershaw T. Text messaging as a tool for behavior change in disease prevention and management. *Epidemiol Rev* 2010 Mar;32:56-69 [FREE Full text] [doi: [10.1093/epirev/mxq004](https://doi.org/10.1093/epirev/mxq004)] [Medline: [20354039](https://pubmed.ncbi.nlm.nih.gov/20354039/)]
49. Fjeldsoe BS, Marshall AL, Miller YD. Behavior change interventions delivered by mobile telephone short-message service. *Am J Prev Med* 2009 Feb;36(2):165-173. [doi: [10.1016/j.amepre.2008.09.040](https://doi.org/10.1016/j.amepre.2008.09.040)] [Medline: [19135907](https://pubmed.ncbi.nlm.nih.gov/19135907/)]
50. Webb TL, Joseph J, Yardley L, Michie S. Using the internet to promote health behavior change: a systematic review and meta-analysis of the impact of theoretical basis, use of behavior change techniques, and mode of delivery on efficacy. *J Med Internet Res* 2010 Feb;12(1):e4 [FREE Full text] [doi: [10.2196/jmir.1376](https://doi.org/10.2196/jmir.1376)] [Medline: [20164043](https://pubmed.ncbi.nlm.nih.gov/20164043/)]
51. U.S. Department of Health and Human Services Office of Minority Health. 2015. Infant Mortality and Hispanic Americans URL: <http://minorityhealth.hhs.gov/omh/browse.aspx?vl=4&vlID=68> [accessed 2016-01-27] [WebCite Cache ID [6esSV6krc](https://www.webcitation.org/6esSV6krc)]

---

## Abbreviations

**MSM:** men who have sex with men

**SMS:** short message service, commonly referred to as “text messaging”

**SNS:** social networking site

---

*Edited by G Eysenbach; submitted 29.01.16; peer-reviewed by L Quintiliani, M Bardus, M Barnes; comments to author 24.03.16; revised version received 06.05.16; accepted 10.06.16; published 11.07.16.*

*Please cite as:*

*Hudnut-Beumler J, Po'e E, Barkin S*

*The Use of Social Media for Health Promotion in Hispanic Populations: A Scoping Systematic Review*

*JMIR Public Health Surveill 2016;2(2):e32*

URL: <http://publichealth.jmir.org/2016/2/e32/>

doi: [10.2196/publichealth.5579](https://doi.org/10.2196/publichealth.5579)

PMID: [27400979](https://pubmed.ncbi.nlm.nih.gov/27400979/)

©Julia Hudnut-Beumler, Eli Po'e, Shari Barkin. Originally published in JMIR Public Health and Surveillance (<http://publichealth.jmir.org>), 11.07.2016. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Public Health and Surveillance, is properly cited. The complete bibliographic information, a link to the original publication on <http://publichealth.jmir.org>, as well as this copyright and license information must be included.

Original Paper

# Testing the Feasibility of a Passive and Active Case Ascertainment System for Multiple Rare Conditions Simultaneously: The Experience in Three US States

Amanda Reichard<sup>1</sup>, PhD; Suzanne McDermott<sup>2</sup>, PhD; Margaret Ruttenber<sup>3</sup>, MSPH; Joshua Mann<sup>4</sup>, MD; Michael G Smith<sup>5</sup>, DrPH; Julie Royer<sup>6</sup>, MSPH; Rodolfo Valdez<sup>7</sup>, MSc, PhD

<sup>1</sup>Institute on Disability, University of New Hampshire, Durham, NH, United States

<sup>2</sup>Department of Epidemiology and Biostatistics, University of South Carolina, Columbia, SC, United States

<sup>3</sup>Special Health Care Needs, Colorado Department of Public Health and Environment, Denver, CO, United States

<sup>4</sup>Department of Preventive Medicine, University of Mississippi Medical Center, Jackson, MS, United States

<sup>5</sup>South Carolina Department of Health and Environmental Control, Columbia, SC, United States

<sup>6</sup>Revenue and Fiscal Affairs Office, South Carolina Budget and Control, Columbia, SC, United States

<sup>7</sup>National Center for Birth Defects and Developmental Disabilities, Centers for Disease Control and Prevention, Atlanta, GA, United States

**Corresponding Author:**

Amanda Reichard, PhD

Institute on Disability

University of New Hampshire

10 West Edge Dr, Suite 101

Durham, NH, 03824

United States

Phone: 1 603 862 5266

Fax: 1 603 862 0555

Email: [amanda.reichard@unh.edu](mailto:amanda.reichard@unh.edu)

## Abstract

**Background:** Owing to their low prevalence, single rare conditions are difficult to monitor through current state passive and active case ascertainment systems. However, such monitoring is important because, as a group, rare conditions have great impact on the health of affected individuals and the well-being of their caregivers. A viable approach could be to conduct passive and active case ascertainment of several rare conditions simultaneously. This is a report about the feasibility of such an approach.

**Objective:** To test the feasibility of a case ascertainment system with passive and active components aimed at monitoring 3 rare conditions simultaneously in 3 states of the United States (Colorado, Kansas, and South Carolina). The 3 conditions are spina bifida, muscular dystrophy, and fragile X syndrome.

**Methods:** Teams from each state evaluated the possibility of using current or modified versions of their local passive and active case ascertainment systems and datasets to monitor the 3 conditions. Together, these teams established the case definitions and selected the variables and the abstraction tools for the active case ascertainment approach. After testing the ability of their local passive and active case ascertainment system to capture all 3 conditions, the next steps were to report the number of cases detected actively and passively for each condition, to list the local barriers against the combined passive and active case ascertainment system, and to describe the experiences in trying to overcome these barriers.

**Results:** During the test period, the team from South Carolina was able to collect data on all 3 conditions simultaneously for all ages. The Colorado team was also able to collect data on all 3 conditions but, because of age restrictions in its passive and active case ascertainment system, it was able to report few cases of fragile X syndrome. The team from Kansas was able to collect data only on spina bifida. For all states, the implementation of an active component of the ascertainment system was problematic. The passive component appears viable with minor modifications.

**Conclusions:** Despite evident barriers, the joint passive and active case ascertainment of rare disorders using modified existing surveillance systems and datasets seems feasible, especially for systems that rely on passive case ascertainment.

(*JMIR Public Health Surveill* 2016;2(2):e151) doi:[10.2196/publichealth.5516](https://doi.org/10.2196/publichealth.5516)

**KEYWORDS**

spina bifida; muscular dystrophy; fragile X syndrome; surveillance

## Introduction

### Methods for Surveillance of Rare Conditions

Surveillance is used to gather data and knowledge that can be used to identify and control a health problem or to improve a public health program or service [1]. Birth defects monitoring programs, which focus primarily on identifying diagnosed cases, have widely used passive systems that gather data from administrative data such as hospital discharge records and administrative records from public insurers [2]. However, these systems can both over- and underestimate the actual prevalence of conditions, because coding for billing purposes is not always accurate or complete [3-5].

Active case ascertainment methods are considered the “gold standard” in public health surveillance [6]. These methods involve trained coders collecting data directly from medical providers, health service providers, and educational institutions to identify cases. The biggest challenge for using these methods is locating people with any of the conditions of interest within a designated geographic area [7]. Thus, population-based approaches must actively review records from diverse sources (eg, inpatient and outpatient settings, rehabilitation services, disability-specific programs, schools) [8].

Rare conditions such as spina bifida, fragile X syndrome, and muscular dystrophy can be especially difficult to monitor. Among these 3 conditions, only spina bifida is recognizable at birth and more easily included in state-based birth defects monitoring systems. The other 2 conditions—fragile X syndrome and muscular dystrophy—are not apparent in the early days of life and no laboratory test or biomarker is commonly used to screen newborns for these conditions. Instead, fragile X syndrome and muscular dystrophy are often identified during early childhood after parents and primary care providers note developmental or motor deficits. Both conditions require specific genetic tests to confirm their diagnosis; however, this usually occurs in an outpatient setting and does not require hospitalization. Thus, the data sources for passive surveillance of these conditions need to be extended not only to genetic laboratories but also to specialty care clinics where diagnosis is confirmed typically after multiple visits [9-11].

### Spina Bifida

Spina bifida occurs when the neural tube fails to close properly during fetal development [12-14]. The incidence of spina bifida detected at birth (namely, meningocele and myelomeningocele) decreased from approximately 2.5/10,000 (95% CI 2.3-2.7) in 1992 to 2.0/10,000 (95% CI 1.9-2.1) in 2001. The National Birth Defects Prevention Network estimated the *prevalence* of spina bifida (without hydrocephaly) was 3.5/10,000 live births (95% CI 3.31-3.68) [14].

### Muscular Dystrophy

Muscular dystrophies are a group of genetic diseases characterized by progressive skeletal muscle degeneration and weakness [15]. Although more than 30 forms of muscular

dystrophy have been identified, there are 9 major forms [16]. The most common muscular dystrophies, Duchenne muscular dystrophy and Becker muscular dystrophy, together have an estimated prevalence of 1.38/10,000 males aged 5 to 24 years [16]. Other major forms (eg, distal, Emery-Dreifuss) each have a prevalence of 1 to 2 per 100,000 [17,18].

### Fragile X Syndrome

Fragile X syndrome results from a mutation in the fragile X mental retardation 1 gene, *FMR1*, on the X chromosome [19]. Impairment severity can range from relatively mild learning disabilities to moderate intellectual disability and autism or “autistic-like” behaviors. Approximately 1 in 3600 to 4000 males and 1 in 4000 to 6000 females is born with the full mutation for fragile X. The vast majority of males and about 50% of females with the full mutation will have fragile X syndrome [19-21].

### Importance of Surveillance of These 3 Rare Conditions

These conditions have low prevalence but a great impact on long-term disability, mortality, cost, and caregiver stress [12,22,23]. Obtaining a better estimate of state prevalence is a necessary starting point for assessing the impact. Such tracking requires the following: a flexible population scope (eg, specific to infants and young children); expansion of data sources (eg, health care specialists and tertiary medical centers); and labor-intensiveness (eg, data abstracted from a large number of health care practices).

Spina bifida, muscular dystrophy, and fragile X syndrome were chosen for this feasibility study because (1) they are all low incidence conditions with high health and economic impact; (2) they encompass a wide range of etiological, physical, and cognitive symptoms; (3) they represent various degrees of difficulty for passive and active case ascertainment; (4) public and private organizations have long-standing active research programs and data collections on these 3 conditions; and (5) the US Congress has provided special funding to monitor and study these 3 conditions [24-27].

The low prevalence of rare disorders makes impractical the development of a separate passive and active case ascertainment system for each condition but lends itself to a combined case ascertainment system that would monitor several conditions at once. Such an integrated passive and active case ascertainment system could serve as a model for other low prevalence and high-impact conditions. Simultaneous passive and active case ascertainment of rare conditions may lead to public health interventions that improve the health of a sizable segment of the population affected by these conditions. Thus, the purpose of this paper is to report on the feasibility of developing a rare conditions passive and active case ascertainment system that simultaneously monitors spina bifida, muscular dystrophy, and fragile X syndrome within a state. In each of the 3 states in which this work was undertaken, objectives were to (1) assess the ability of the local passive and active case ascertainment systems to capture the 3 rare disorders, (2) capture preliminary



state prevalence estimates of the conditions, and (3) discover barriers and facilitators to such implementation. Findings can provide lessons for future rare conditions passive and active case ascertainment activities in states with similar systems to these.

## Methods

### Justification of States Included in the Study

The design of this feasibility study required that the states included were at varying levels of readiness toward implementing a rare conditions passive and active case ascertainment program. The state teams were selected by the study team, based on knowledge of states with disability epidemiologists and their understanding of data capability. States were also selected so that they roughly conformed to one each on the high, medium, and lower levels of passive and active case ascertainment sophistication. Colorado (high) and South Carolina (medium) were selected because of their established and state-supported birth defects passive and active case ascertainment system and, in the case of South Carolina, other integrated data system capacities. Kansas (low) was selected because it was home to a disability epidemiologist who had published articles using the Medicaid system to analyze disability and health issues. The approach in each state was to identify existing passive and active case ascertainment systems and data sources that could be expanded to implement a more comprehensive system within 2 years. The 2-year duration was chosen to allow up to 1 year for the assessment of current passive and active case ascertainment systems, standardization of case definitions, and selection of variables for the active case ascertainment component; and an additional year to test the feasibility of modestly enhancing the existing systems to assess what is feasible and to compare the three systems. Within each state, systems were compared by (1) using data from the active system to assess the accuracy of cases identified in the passive system and (2) capture-recapture methodology to get estimates of the prevalence of the 3 conditions.

The first step was to document how states varied in their different resources, approaches, and levels of experience in conducting passive and active case ascertainment. [Table 1](#) compares passive and active case ascertainment programs in the 3 states [28,29]. The same sources and variables were searched and assessed for all 3 conditions.

Colorado has a mature, state-mandated, birth defects surveillance system in the health department, which requires reporting up to age 3 years. This system conducts active case finding through administrative data sources and it includes reports of spina bifida that mirror expected prevalence. Colorado also has a Centers for Disease Control and Prevention (CDC)-funded muscular dystrophy surveillance program (Muscular Dystrophy Surveillance, Tracking, and Research Network; MD STARnet) that uses active surveillance in neuromuscular clinics to identify childhood-onset cases. The MD STARnet system had been ongoing for 10 years at the beginning of our project. The

challenge for Colorado was to identify cases of fragile X syndrome, because there had only been a few cases reported in the past decade. Fragile X syndrome is an inherited cause of intellectual disability, which is not apparent at birth. When intellectual disability is diagnosed in a child, a genetic diagnostic examination is needed to make the fragile x syndrome diagnosis; thus, in many cases this is not done before the child's third birthday. The fact that Colorado only had a few cases of fragile X syndrome reported before the onset of the project suggests that even established birth defects systems need expanded authority through their legislature to collect information about people who are older than 3 years. The Colorado passive and active case ascertainment system needs to pursue sources of reporting, such as the fragile X clinic in their state, if they want to conduct passive and active case ascertainment of rare conditions. South Carolina has a birth defects case ascertainment system that relies on active hospital record abstraction for spina bifida and other early identifiable birth defects. This system had been collecting cases of spina bifida at the prevalence rate that was expected. South Carolina also has a well-established administrative data system that allows linkage among a large number of public insurance, vital records, and state service agency secondary data sources. This system was available for passive ascertainment. The South Carolina project developed an active surveillance system in 5 counties to compare the active and passive ascertainment of the 3 conditions. Kansas, likewise, has a birth defects surveillance system housed in the health department that primarily relied on birth certificate reporting of birth defects. This system has a passive case ascertainment component and no active case ascertainment component. Thus, the challenge for Kansas was to develop a small, active, case ascertainment component for this project.

During the first year of this 2-year project the teams from Colorado and South Carolina began with the identification of existing data sources. The Kansas team spent the first year cataloging features they could use from the other 2 states that had well-developed passive and active case ascertainment systems, investigating the capacity of local chapters of parent advocacy organizations for each condition, and conducting an extensive literature review. The Kansas team worked with the South Carolina and Colorado health departments to obtain information about policies and procedures related to existing passive and active case ascertainment systems, particularly on birth defects monitoring programs.

Together, the teams from the 3 states established uniform case definitions (including International Classification of Diseases, Ninth Revision, or ICD-9 codes and other required elements), variable definitions, and data abstraction tools for the active case ascertainment approach. The codes used in the 3 states are listed in [Table 2](#). South Carolina and Kansas completed institutional review board (IRB) reviews that were needed because an active case ascertainment system that covered all ages was being used for this project, and this expanded an existing state statutory system.

**Table 1.** Major characteristics of the case ascertainment system in 3 US states.

Characteristics	Colorado	South Carolina	Kansas
<b>State characteristics</b>			
Legislative authority	CRS <sup>a</sup> 25-1.5-101 to 25-1.5-105; enacted in 1985	A281,R308,H4115; enacted in 2004	KSA <sup>b</sup> 65-1241 to 65-1246; enacted in 2004
Location	Department of Health: Epidemiology and Environment	Department of Health: Maternal and Child Health	Department of Health: Vital Statistics, Maternal and Child Health
Characterization of the passive and active case ascertainment system	Mostly passive	Passive and active	Passive only
<b>Data characteristics</b>			
Data sources	Case ascertainment (active, passive); vital records (birth, death, and fetal death certificates); state-based registries; delivery hospitals; pediatric and tertiary care hospitals; genetic laboratories; genetic counseling services; genetic clinics; physician reports.	Case ascertainment (active); vital records (birth, death, fetal death, and elective termination certificates); state-based registries; delivery hospitals; pediatric and tertiary hospitals; prenatal diagnosis facilities; genetic laboratories; genetic counseling services; genetic clinics; physician reports; passive sources; Medicaid; hospital discharges; state health plan claims; Department of Disabilities and Special Needs.	Case ascertainment (passive); vital records (birth and fetal death certificates); state-based registries; physician reports.
Time frame—years covered for passive case ascertainment	SB <sup>c</sup> : 2010-12 MD <sup>d</sup> : 1992-2011 FXS <sup>e</sup> : 2007-12	SB: 1996-2012 MD: 1996-2012 FXS: 1996-2012 (secondary data with code)	SB: 1979-2013
Time frame—years covered for active case ascertainment	SB: 2010-12 MD: 1992-2011 FXS: 2007-12	SB: 2013-2014 MD: 2013-2014 FXS: 2013-2014 (for prevalent cases)	SB: 1971-2013
Age range covered, years	0-3 (SB, FXS); 0-28 (MD)	No limit	No limit
Data sources for the active case ascertainment component in this project	Reascertainment and medical records.	Medical records (5 counties).	Medical records (1 county).
Clinical review for the passive component of this project	Yes	No	No
Clinical review for the active component of this project	Yes	Yes	Yes
Barriers to implementation	Not enough time to change the age of reporting FXS from 3 to 10 years.	The active component could be labor intensive.	Limited availability and interconnection of data sources.

<sup>a</sup>CRS: Colorado Revised Statutes.

<sup>b</sup>KSA: Kansas Statutes Annotated.

<sup>c</sup>SB: spina bifida.

<sup>d</sup>MD: muscular dystrophy.

<sup>e</sup>FXS: fragile X syndrome.



**Table 2.** Diagnosis codes and variables used to passively identify cases in a pilot project for a 3-state (Colorado, Kansas, South Carolina) public health passive and active case ascertainment system for 3 rare conditions (spina bifida, muscular dystrophy, and fragile X syndrome).

Condition	Hospital discharge and insurance ICD-9-CM <sup>a</sup> all diagnosis fields	Birth certificates Source-specific indicator	Death certificates ICD-10 <sup>b</sup> all cause fields
Fragile X syndrome	759.83	N/A <sup>c</sup>	Q99.2
Muscular dystrophy	359.0, 359.1, and 359.21	N/A	G71.0 and G71.1
Spina bifida (without anencephalies)	741.0 through 741.9	Myelomeningocele or meningocele variable	Q05 including Q05.0 to Q05.9

<sup>a</sup>ICD-9-CM: International Classification of Diseases, Ninth Revision, Clinical Modification.

<sup>b</sup>ICD-10: International Classification of Diseases, Tenth Revision.

<sup>c</sup>N/A: not applicable.

The teams assessed the status of statewide data sources for rare conditions passive and active case ascertainment in their respective states. All 3 states have birth defects registries that obtain data from birth certificates. Colorado and South Carolina queried electronic administrative data sources, including hospital and emergency department encounters, physician office visits, all of which include codes for conditions, services, and charges, as well as death records.

During the second year the states carried out both active and passive case ascertainment, documented barriers and challenges as they arose, and made tentative estimations of the prevalence of spina bifida, muscular dystrophy, and fragile X syndrome in each state.

### Methods by State

Each state's methodology was based on its readiness for combined passive and active case ascertainment, the organization that led the project, the type of available data, and the answers to unanticipated barriers to the implementation of the passive and active case ascertainment system. The Guidelines for Conducting Birth Defects Surveillance [24] was the document that set the standards for operations in both Colorado and South Carolina and it was used to establish activities during this feasibility project.

#### Colorado

The Colorado Department of Public Health and Environment has a sophisticated, mainly passive, birth defects case ascertainment system that interfaces with some administrative data sources. The state has more than 20 years of surveillance data collection, under statutory authority, on a large array of childhood conditions, which include active case ascertainment data for muscular dystrophy and spina bifida.

#### Colorado's Approach to Passive Case Ascertainment for Persons With Fragile X Syndrome, Spina Bifida, and Muscular Dystrophy

Colorado Responds to Children with Special Needs (CRCSN) is the program in charge of birth defects monitoring and prevention in the state. This program used case reports from multiple external sources to ascertain cases of each of the 3 conditions of interest among children from birth to age 3 years, for all cases meeting criteria. Case reports were entered into a

transitional computer program that prepared the case for further processing before being posted to a core database. All case reports went through an extensive search and/or match process that linked cases to a unique identifier.

Colorado has participated in a number of CDC-funded surveillance efforts that have enhanced its data collection processes. For example, CRCSN collected data on children with spina bifida up to age 3 years through the CDC-funded Rapid Ascertainment project that uses a passive registry followed by a reascertainment and medical record review as confirmation for all live births [30]. Colorado has also participated in the CDC-funded MD STARnet program since it began surveillance for Duchenne and Becker muscular dystrophies in 2002 [31]. Recently, MD STARnet was expanded to include 7 additional types of muscular dystrophy. CRCSN has collected information on fragile X syndrome since 1994, through its passive Birth Defects Monitoring system, but only a few cases were reported in recent years. As part of this feasibility passive and active case ascertainment effort the Colorado project staff met with the director of the fragile X clinic at Children's Hospital Colorado in an effort to expand the passive ascertainment and explore the possibility of conducting a medical record review to identify persons for whom a genetics laboratory confirmation was available. This contact was necessary to inform the director of the clinic's responsibility to report children younger than 3 years with fragile X syndrome to the CRCSN system.

#### Colorado's Approach to Active Case Ascertainment of Persons With Fragile X Syndrome, Spina Bifida, and Muscular Dystrophy

The CRCSN staff established specific guidelines for each case definition, which required the number of times a diagnosis is reported, the number of reporting sources, and the use of International Classification of Diseases, Ninth Revision, Clinical Modification (ICD-9-CM) diagnosis and procedure codes. The CRCSN did not implement an active case ascertainment system per se; instead, it supplemented the passive system with rigorous reascertainments and medical record reviews by physicians of all spina bifida cases and some muscular dystrophy cases. Before this feasibility project, this approach was not used for cases of fragile X syndrome.

CRCSN has developed a multistep approach to monitor data sources, specific diagnoses, over- and under-ascertainment, and

problematic code assignment issues. Its data quality validation procedure includes computerized and manual approaches (computer screen review of case records, medical records review of selected cases, and clinical consultation when difficulties were identified). Additionally, two levels of clinical reviews are used for diagnosis: (1) staff level (to identify problems in report date, source of diagnosis, site of the encounter, and ICD-9-CM codes); (2) medical specialist level (to review the medical records and other relevant information to verify the diagnosis).

### **South Carolina**

In this state, faculty from the University of South Carolina School of Public Health and the Medical University of South Carolina partnered with the South Carolina Department of Health and Environmental Control (DHEC) to complete this project. The active case ascertainment was conducted by the DHEC and the data utilized for the passive case ascertainment system was housed within the Revenue and Fiscal Affairs (RFA) agency. The RFA has agreements with state agencies and organizations to store data, although each data source retains control of the data at all times. Thus, RFA facilitated the development of written agreements with each agency and organization that potentially had data elements that could be used for this project. For deidentified data projects, including this rare conditions passive and active case ascertainment effort, the RFA uses an algorithm that relies on source-specific personal identifiers to create a global unique identifier. The data are edited and standardized before being run through the search algorithm. The global unique identifier is not associated with any personal identifier and is used on all subsequent episodes of services, regardless of data source or service provider. Using the unique global identifier enables staff to use data across multiple providers while protecting confidentiality. The investigators of the rare conditions passive and active case ascertainment system included one RFA staff member, a health department investigator, and university investigators.

### **South Carolina's Approach to Passive Case Ascertainment of Persons With Fragile X Syndrome, Spina Bifida, and Muscular Dystrophy**

All persons with an ICD-9-CM primary or secondary diagnosis code for fragile X syndrome, spina bifida, or muscular dystrophy were identified from Medicaid, the Hospital Discharge Dataset, and the claims from the State Health Plan that insures all government workers and their families (Table 2). The South Carolina DHEC Birth Defects Monitoring Program was used to identify cases of spina bifida. For all 3 conditions, death certificate records were used to identify unique cases from previous periods. These cases occur if people received care out of state or if they did not have their condition diagnosis included when they were seen in the medical care system. These cases were identified by International Classification of Diseases, Tenth Revision, (ICD-10) cause-of-death codes for fragile X syndrome, spina bifida, and muscular dystrophy from the South Carolina DHEC. For fragile X syndrome and muscular dystrophy, South Carolina also identified cases through the Department of Disabilities and Special Needs, the agency that provides support and services to people with disability.

### **South Carolina's Approach to Active Case Ascertainment of Persons With Fragile X Syndrome, Spina Bifida, and Muscular Dystrophy**

South Carolina DHEC established an active case ascertainment component for citizens of all ages with rare conditions in 5 South Carolina counties for this project, after IRB approval. After cases were reported to the South Carolina DHEC Birth Defects Monitoring Program by the physician practices, a registered nurse abstractor traveled to each practice to abstract relevant data (ie, basic diagnosis and demographic information; sufficient but less rigorous than the Colorado MD STARnet protocol for muscular dystrophy) from medical records of those who had received treatment in a clinic, hospital, or practice located in those 5 counties. The data were then entered into an Epi Info [32] data system for compilation, output into an SAS [33] file for editing, linking with passive data, and analysis.

### **South Carolina's Approach to Merging the Passive and Active Systems**

At the conclusion of active data collection for the 5 designated counties, the data from the passive case ascertainment system and the active case ascertainment system were linked and analyzed at RFA. The data collected through the passive system were compared with the data from the active case ascertainment system to test the accuracy and completeness of the passive system. Non-RFA investigators were provided with aggregate reports.

### **Kansas**

Congenital anomalies have been recorded in Kansas birth certificates since 1979 and since 1982, Kansas has had, under administrative regulations, a limited set of passive case ascertainment activities for these anomalies [28]. In 2004, the Kansas legislature issued the statutes for the creation of the Birth Defects Information System (BDIS) with the aim of collecting information on congenital anomalies and other birth abnormalities among children younger than 5 years [28]. Currently, the BDIS includes an interface between a birth defects database and a Web-based application from the program Children and Youth with Special Health Care Needs, along with relevant variables from the Vital Statistics Integrated Information System. Thirteen anomalies are currently listed in the birth certificates and reported to the BDIS [28].

### **Kansas' Approach to Passive Case Ascertainment for Persons With Fragile X Syndrome, Spina Bifida, and Muscular Dystrophy**

Given the strong reliance of the BDIS on birth certificates and because spina bifida is the only one of the 3 conditions included in this report that is diagnosed at birth, the Kansas team developed a passive case ascertainment plan just for spina bifida. First, the team identified cases of spina bifida using ICD-9-CM codes (see Table 2) in a Medicaid claims database. Then, relevant individual information from these cases was merged with individual information from the BDIS. Finally, the individual data were deduplicated and aggregated to obtain the count of spina bifida cases in the state.

### **Kansas' Approach to Active Case Ascertainment for Persons With Fragile X Syndrome, Spina Bifida, and Muscular Dystrophy**

For the active case ascertainment component, the Kansas team mailed a letter to pediatricians, neurologists, and spina bifida clinics in one county (Sedgwick) in Kansas asking them to report all of their cases of spina bifida, under authority of the Kansas Birth Defects Act (KSA 65-1241 through 65-1246). Sedgwick County, including the city of Wichita, is a large county, including 17.5% of the state's total population [34]. For professional medical record abstraction, eligible medical records were sent to the Kansas Foundation for Medical Care (KFMC), a state not-for-profit organization that has served as the state's External Quality Review Organization and information technology resource in the state. Additionally, KFMC provided advice on data aggregation.

### **Kansas' Approach to Merging the Passive and the Active Case Ascertainment Systems**

When the active component of the project was completed, KFMC shared the active case ascertainment dataset with the Kansas team. These data were merged with the passive dataset to identify the cases that were common between both datasets.

## **Results**

### **Findings by State**

Within the 2-year period allocated to this project, the South Carolina team was able to use the passive and active case ascertainment capabilities of the state to assemble a passive and active case ascertainment system, with both passive and active components, to simultaneously detect spina bifida, muscular dystrophy, and fragile X syndrome. The active component of the South Carolina team, however, was limited to 5 counties. The Colorado team supplemented their extensive passive system with reascertainment and clinical review of the cases; however, because of age restrictions in reporting, this team was able to detect only a handful of cases of fragile X syndrome. The Kansas team was able to assemble a case ascertainment system with passive and active components, but only for 1 condition, spina bifida, and the active component of this system was limited to 1 county.

On the basis of cases found, the teams calculated number and prevalence for their respective states. These estimates are tentative, not comparable in any way from one state to the other, and certainly not comparable to published estimates calculated with more rigorous methods. Table 3 presents the data collected in the 3 states, collected for spina bifida, muscular dystrophy, and fragile X syndrome. The racial composition of cases in the 3 states differed substantially, but this composition follows the underlying population of each state. In 2014, a total of 66% of the South Carolina population was white, 27% black, and 5% Hispanic, and the cases of spina bifida and muscular dystrophy reflect this pattern. The cases of fragile X syndrome are 47% white for active and 54% for passive. In Colorado and Kansas, there were high proportions of missing data or "other" racial groups noted. The male to female ratio for spina bifida was not consistent across the 3 areas. More females than males were identified in South Carolina and more males than females were identified in both Colorado and Kansas. The higher proportion of male cases identified with muscular dystrophy and fragile X syndrome in Colorado and South Carolina (muscular dystrophy only) reflects the fact that these 2 conditions have genetic X-linked inheritance.

Table 4 presents the rates of spina bifida, muscular dystrophy, and fragile X syndrome identified in this pilot project. In Colorado, the rate for spina bifida was 3.35/10,000 for those younger than 3 years, for muscular dystrophy the rate was 1.29/10,000, and with only 6 identified cases of fragile X syndrome the rate was close to zero. The passive and active case ascertainment efforts in South Carolina were for all ages, so the denominator for South Carolina was the state population for passive case ascertainment and the all age population of 5 counties around the metropolitan areas of Columbia and Charleston for active case ascertainment. The South Carolina passive rates were substantially higher than active rates for spina bifida (2.16/10,000 for active and 12.15/10,000 for passive) and muscular dystrophy (3.28/10,000 for active and 6.84/10,000 for passive). For fragile X syndrome these rates were more similar, 1.20/10,000 for active and 1.65/10,000 for passive case ascertainment. In Kansas, for the one county used for active case ascertainment of spina bifida the rate was 0.86/10,000 and the passive rate was 3.04/10,000.

**Table 3.** Summary of data collected in a pilot project for a 3-state (Colorado, Kansas, South Carolina) public health passive and active case ascertainment system for 3 rare conditions (spina bifida, muscular dystrophy, and fragile X syndrome).

Condition	Colorado <sup>a</sup> (64 counties) n (%)	South Carolina ac- tive (5 counties) n (%)	South Carolina passive (46 counties) n (%)	Kansas active (1 county) n (%)	Kansas passive (statewide) n (%)
Spina bifida, N	... <sup>b</sup>	253	5872	...	882
Non-Hispanic black	...	73 (28.8)	1590 (27.1)	...	25 (2.8)
Non-Hispanic white	30 (...)	138 (54.5)	3628 (61.8)	21 (...)	296 (33.6)
Hispanic and others	28 (...)	42 (16.6)	654 (11.1)	19 (...)	561 (63.6)
Male	36 (61.0)	112 (44.3)	2266 (38.6)	26 (59.1)	392 (55.6)
Female	10 (16.9)	141 (55.7)	3606 (61.4)	18 (40.9)	490 (44.4)
Missing	13 (22.0)				
Birth year, range and 95% CI	Range 2010-2012	95% CI 1982-2014	95% CI 1958-1998	Range 1971-2013	Range 2004-2013
Muscular dystrophy <sup>c</sup> , N	689	384	3305		
Non-Hispanic black	16 (2.3)	75 (19.5)	805 (24.4)		
Non-Hispanic white	379 (55.0)	262 (68.2)	1998 (60.5)		
Hispanic and others	294 (42.7)	47 (12.2)	502 (15.2)		
Male	444 (64.4)	258 (67.2)	1707 (51.6)		
Female	245 (35.6)	126 (32.8)	1598 (48.4)		
Birth year, range and mean (SD)	1992-2011	1973 (23)	1969 (24)		
Fragile X syndrome, N	6	141	795		
Non-Hispanic black		57 (40.4)	286 (36.0)		
Non-Hispanic white		66 (46.8)	433 (54.5)		
Hispanic and others		18 (12.7)	76 (9.6)		
Male		88 (62.4)	516 (64.9)		
Female		53 (37.6)	279 (35.1)		
Birth year, range and mean (SD)	2007-2012	1983 (19)	1981 (19)		

<sup>a</sup>Colorado did not implement an active case ascertainment system per se; instead, it supplemented the passive system with reascertainments and medical record reviews of all spina bifida cases and some muscular dystrophy cases.

<sup>b</sup>Ellipses indicate that the cells contain less than 5 individuals; owing to confidentiality concerns, the exact number has been suppressed.

<sup>c</sup>South Carolina and Colorado differed in the rigor of the active case ascertainment. Colorado, as part of Centers for Disease Control and Prevention's Muscular Dystrophy Surveillance, Tracking, and Research Network (MD STARnet), used a very thorough protocol for active case ascertainment. South Carolina was less intensive, recording fewer key variables. Both recorded counts for the 9 major forms of muscular dystrophy. Kansas did not collect data on Muscular Dystrophy or Fragile X.



**Table 4.** Prevalence rates of spina bifida, muscular dystrophy, and fragile X syndrome in the 3 states, based on reference population.

Measure	South Carolina	Colorado	Kansas
State and subarea population used for rate calculation	4,832,482 state population for passive	176,169 state for younger than 3 years	2,904,021 state population for passive
(2014)	1,170,249 for 5 counties <sup>a</sup> for active case ascertainment		508,803 for 1 county for active case ascertainment
Rate of spina bifida	2.16/10,000 active 12.15/10,000 passive	3.35/10,000 younger than 3 years	0.86/10,000 active <sup>b</sup> 3.04/10,000 passive
Rate of muscular dystrophy	3.28/10,000 active 6.84/10,000 passive	1.29/10,000	N/A <sup>c</sup>
Rate of fragile X syndrome	1.20/10,000 active 1.65/10,000 passive	0.01/10,000	N/A

<sup>a</sup> The 5 South Carolina counties for active case ascertainment are as follows: Berkeley, Dorchester, Charleston, Lexington, and Richland.

<sup>b</sup> Only 1 Kansas county was included in active case ascertainment.

<sup>c</sup> N/A: not applicable.

## Barriers to Implementation

The Colorado team reported no major problems implementing this project as they already actively ascertained or reascertained from passive reports the cases of muscular dystrophy and spina bifida. The limited time for this project (2 years) did not allow Colorado to officially request approval from the state legislature to change the age of reporting for fragile X syndrome from 3 to 10 years, as has been done with autism and fetal alcohol syndrome. Therefore, Colorado did not have any prevalent cases of fragile X syndrome to report and only 6 incident cases.

The major barrier to implementing a combined passive and active case ascertainment system for rare disorders in South Carolina is cost. Passive data sources are readily available and the active case ascertainment system is a natural extension of the ongoing DHEC Birth Defects Monitoring Program; however, the incremental cost of expanding active case ascertainment can be substantial. For this project, a full-time registered nurse with previous active case ascertainment experience was hired. The costs for the project included her salary and fringe benefits, training and travel, the customary clinical review of cases found, and the storage and protection of data. Finally, the utility of identifying and monitoring new cases needs to be justified with clear benefits for the patients and their caregivers.

In Kansas, the BDIS relies heavily on data from birth certificates with the age of 5 years being the limit for reporting genetic or congenital conditions [28,29]. The monitoring of conditions that are detectable long after birth under such constraints, such as fragile X syndrome and muscular dystrophy, could be challenging. After limiting the test to spina bifida, a major policy barrier surfaced. The Kansas law that created BDIS also made the records contained in this system strictly confidential and disclosable only by court order. Public disclosure is only allowed in summary or aggregated formats (law identified in Table 1); therefore, merging, matching, and analyzing the data had to be performed in situ by team members affiliated with the Kansas Department of Health and Environment.

To facilitate data sharing regarding medical records, the Kansas team contacted KFMC to perform medical record abstraction

for this project. The KFMC has an established and credible relationship with health care providers in Kansas and recently partnered with the Kansas Department of Health and Environment to complete a record review project on their behalf. However, even with this established relationship and the state law, there was a limited response from providers to participate in the active component.

## Discussion

This passive and active case ascertainment project was carried out to test the feasibility of establishing rare conditions passive and active case ascertainment systems for more than one condition at a time in states with varying levels of existing infrastructure. Having states at various stages of readiness for passive and active case ascertainment in this project allows for the identification of factors that may facilitate or impede the development of such systems.

The approach was to first assess the existing birth defects monitoring system in all 3 states, and then to identify additional processes that could be used to implement an ongoing rare conditions passive and active case ascertainment system. This project was not designed to compare active with passive case ascertainment; rather, it was designed to assess the feasibility of combining both types of case ascertainment systems to increase the yield of cases. The early steps for this project were successful: separately, all teams from the 3 states identified local data sources and determined the data elements to be collected. Together, the teams worked on the standardization of variable definitions to make the results comparable across systems and states. The definitions of the variables and their connections to the data sources were precisely documented to assure accurate replication.

This feasibility exercise helped us better understand the ways that states approach passive and active case ascertainment. Some states, like South Carolina and Colorado, have legal authority and ample experience monitoring birth defects. These 2 states use a variety of data sources for their passive and active case ascertainment systems, but active case ascertainment of rare diseases relies on obtaining funding through national programs,

such as MD STARnet. In fact, after completion of this project, South Carolina was able to apply and get funding to join the national MD STARnet network. On the other hand, for other states, such as Kansas, the birth defects monitoring may have limited data sources and the implementation of a passive and active case ascertainment system to monitor more than one rare condition at once with both passive and active components would be a major challenge. Thus, as has been seen with other passive and active case ascertainment systems, funding through national networks or advocacy foundations appears to be the most viable approach to support the establishment of passive and active case ascertainment of rare conditions [35].

Because the purpose of the project was to assess the feasibility of establishing active and passive case ascertainment for rare conditions, we cannot guarantee the accuracy of the prevalence estimates; we have presented them here to provide reference for future work. Although the rates are not directly comparable because of state-to-state variations in data sources, prevalence of spina bifida was 2.16/10,000 people in South Carolina (active case ascertainment component), 3.35/10,000 in Colorado (passive case ascertainment component supplemented with case confirmations and clinical reviews), and 3.04/10,000 in Kansas

(passive case ascertainment component; active case ascertainment resulted in limited follow-up). The prevalence of muscular dystrophy differed between South Carolina (3.28/10,000) and Colorado (1.29/10,000), although South Carolina used less rigorous criteria than the MD STARnet criteria and process [32] used by Colorado staff to define 9 forms of the condition. South Carolina was the only state able to test a combined passive and active case ascertainment approach for fragile X syndrome; the prevalence was 1.20/10,000 for active case ascertainment and 1.65/10,000 for passive case ascertainment.

The lessons learned have been valuable for the 3 states participating in this pilot project. Team participation and problem-solving approaches were excellent; however, state health departments will face obstacles if they seek to implement a combined passive and active case ascertainment system to simultaneously track rare disorders statewide. Although joint passive and active case ascertainment of rare disorders is feasible, only the passive component of case ascertainment seems to be readily available for use, even with limitations regarding data collection. The incorporation of the active component appears to be more of a long-term goal.

---

## Acknowledgments

The authors would like to thank the following staff members of the Kansas Department of Health and Environment for their collaboration and support throughout this project: Greg Crawford, Joy Crevoiserat, Jamie Kim, and Heather Smith.

This research was supported by a cooperative agreement U01-DD000927 with the Centers for Disease Control and Prevention's National Center on Birth Defects and Developmental Disabilities. The findings and conclusions of this report are those of the authors and do not necessarily represent the views of the Centers for Disease Control and Prevention, the South Carolina Revenue and Fiscal Affairs Office, the South Carolina Department of Health and Human Services, the South Carolina Public Employee Benefit Authority, the South Carolina Department of Disabilities and Special Needs, South Carolina Department of Health and Environmental Control, or the other agencies that participated in state-based passive and active case ascertainment.

---

## Conflicts of Interest

None declared.

---

## References

1. Hall H, Correa A, Yoon P, Braden C, Centers for Disease Control and Prevention. Lexicon, definitions, and conceptual framework for public health surveillance. *MMWR Suppl* 2012 Jul 27;61(3):10-14. [Medline: [22832991](#)]
2. Mai C, Kirby R, Correa A, Rosenberg D, Petros M, Fagen M. Public Health Practice of Population-Based Birth Defects Surveillance Programs in the United States. *J Public Health Manag Pract* 2016;22(3):E1-E8. [doi: [10.1097/PHH.0000000000000221](#)] [Medline: [25905668](#)]
3. Salemi J, Tanner J, Sampat D, Anjohrin S, Correia J, Watkins S, et al. The Accuracy of Hospital Discharge Diagnosis Codes for Major Birth Defects: Evaluation of a Statewide Registry With Passive Case Ascertainment. *J Public Health Manag Pract* 2016;22(3):E9-E19. [doi: [10.1097/PHH.0000000000000291](#)] [Medline: [26125231](#)]
4. Schoellhorn J, Collins S. False positive reporting of Hirschsprung's disease in Alaska: an evaluation of Hirschsprung's disease surveillance, birth years 1996-2007. *Birth Defects Res A Clin Mol Teratol* 2009 Nov;85(11):914-919. [doi: [10.1002/bdra.20628](#)] [Medline: [19777544](#)]
5. Fox DJ, Druschel CM. Estimating prevalence of fetal alcohol syndrome (FAS): effectiveness of a passive birth defects registry system. *Birth Defects Res A Clin Mol Teratol* 2003 Sep;67(9):604-608. [doi: [10.1002/bdra.10108](#)] [Medline: [14703781](#)]
6. Honein MA, Paulozzi LJ. Birth defects surveillance: assessing the "gold standard". *Am J Public Health* 1999 Aug;89(8):1238-1240. [Medline: [10432914](#)]
7. Nsubuga P, White ME, Thacker SB, Anderson MA, Blount SB, Broome CV, et al. Public Health Surveillance: A Tool for Targeting and Monitoring Interventions. In: Jamison DT, Breman JG, Measham AR, editors. *Disease Control Priorities in*

- Developing Countries. 2nd edition. Washington, DC: The International Bank for Reconstruction and Development/The World Bank; 2006.
8. Centers for Disease Control and Prevention. Enhancing Surveillance. In: Manual for the surveillance of vaccine-preventable diseases. 5th edition. Atlanta: Centers for Disease Control Prevention; 2011.
  9. Centers for Disease Control Prevention (CDC). Prevalence of Duchenne/Becker muscular dystrophy among males aged 5-24 years - four states, 2007. *MMWR Morb Mortal Wkly Rep* 2009 Oct 16;58(40):1119-1122 [[FREE Full text](#)] [Medline: [19834452](#)]
  10. Bailey DB, Raspa M, Bishop E, Holiday D. No change in the age of diagnosis for fragile x syndrome: findings from a national parent survey. *Pediatrics* 2009 Aug;124(2):527-533. [doi: [10.1542/peds.2008-2992](#)] [Medline: [19581269](#)]
  11. Salemi JL, Tanner JP, Kennedy S, Block S, Bailey M, Correia JA, et al. A comparison of two surveillance strategies for selected birth defects in Florida. *Public Health Rep* 2012;127(4):391-400 [[FREE Full text](#)] [Medline: [22753982](#)]
  12. Sandler AD. Children with spina bifida: key clinical issues. *Pediatr Clin North Am* 2010 Aug;57(4):879-892. [doi: [10.1016/j.pcl.2010.07.009](#)] [Medline: [20883878](#)]
  13. Mathews TJ, Honein MA, Erickson JD. Spina bifida and anencephaly prevalence--United States, 1991-2001. *MMWR Recomm Rep* 2002 Sep 13;51(RR-13):9-11 [[FREE Full text](#)] [Medline: [12353510](#)]
  14. Parker S, Mai CT, Canfield MA, Rickard R, Wang Y, Meyer RE, National Birth Defects Prevention Network. Updated National Birth Prevalence estimates for selected birth defects in the United States, 2004-2006. *Birth Defects Res A Clin Mol Teratol* 2010 Dec;88(12):1008-1016. [doi: [10.1002/bdra.20735](#)] [Medline: [20878909](#)]
  15. Emery AE. The muscular dystrophies. *Lancet* 2002 Feb 23;359(9307):687-695. [doi: [10.1016/S0140-6736\(02\)07815-7](#)] [Medline: [11879882](#)]
  16. National Institute of Child Health and Human Development. 2013. What are the types of muscular dystrophy? URL: <https://www.nichd.nih.gov/health/topics/musculardys/conditioninfo/Pages/types.aspx> [accessed 2016-01-11] [[WebCite Cache ID 6eSSUcpVU](#)]
  17. Romitti P, Zhu Y, Puzhankara S, James K, Nabukera S, Zamba G, et al. Prevalence of Duchenne and Becker Muscular Dystrophies in the United States. *PEDIATRICS* 2015;135(3):513-521.
  18. Centers for Disease Control and Prevention. Facts about Muscular Dystrophy URL: <http://www.cdc.gov/ncbddd/muscular dystrophy/facts.html> [accessed 2016-01-11] [[WebCite Cache ID 6eSSp5yjB](#)]
  19. Crawford DC, Acuña JM, Sherman SL. FMR1 and the fragile X syndrome: human genome epidemiology review. *Genet Med* 2001;3(5):359-371 [[FREE Full text](#)] [Medline: [11545690](#)]
  20. Hagerman RJ, Hagerman PJ. The fragile X premutation: into the phenotypic fold. *Curr Opin Genet Dev* 2002 Jun;12(3):278-283. [Medline: [12076670](#)]
  21. Saul RA, Friez M, Eaves K, Stapleton GA, Collins JS, Schwartz CE, et al. Fragile X syndrome detection in newborns-pilot study. *Genet Med* 2008 Oct;10(10):714-719. [doi: [10.1097/GIM.0b013e3181862a76](#)] [Medline: [18813135](#)]
  22. Iosif A, Sciolla AF, Brahmabhatt K, Seritan AL. Caregiver Burden in Fragile X Families. *Curr Psychiatry Rev* 2013 Feb 1;9(1) [[FREE Full text](#)] [doi: [10.2174/157340013805289590](#)] [Medline: [24348279](#)]
  23. Ouyang L, Grosse S, Riley C, Bolen J, Bishop E, Raspa M, et al. A comparison of family financial and employment impacts of fragile X syndrome, autism spectrum disorders, and intellectual disability. *Res Dev Disabil* 2014 Jul;35(7):1518-1527 [[FREE Full text](#)] [doi: [10.1016/j.ridd.2014.04.009](#)] [Medline: [24755230](#)]
  24. National Fragile X Foundation. FragileX. 2016. National Fragile X Foundation URL: <https://fragilex.org/research> [accessed 2016-06-15] [[WebCite Cache ID 6jmdYqS07](#)]
  25. Muscular DA. mda. 2016. MDA: For Strength Independence & Life URL: <https://www.mda.org/> [accessed 2016-08-15] [[WebCite Cache ID 6jmdhCwZM](#)]
  26. Parent Project Muscular Dystrophy. parentprojectmd. 2016. Parent Project Muscular Dystrophy: Leading the Fight to End Duchenne URL: <http://www.parentprojectmd.org/site/PageServer?pagename=Understand> [accessed 2016-06-15] [[WebCite Cache ID 6jmeleYVt](#)]
  27. Spina Bifida Association of America. Spina Bifida Association. 2016. Spina Bifida Association of America URL: <http://spinabifidaassociation.org/> [accessed 2016-08-15] [[WebCite Cache ID 6jmeBrnAa](#)]
  28. National Center on Birth Defects Developmental Disabilities (CDC). State Birth Defects Surveillance Program Directory. *Birth Defects Research (Part A): Clinical and Molecular Teratology* 2011;91:1028-1149. [Medline: [17109399](#)]
  29. National Birth Defects Prevention Network. National Birth Defects Prevention Network. National Birth Defects Prevention Network: Birth Defects Surveillance, Research & Prevention URL: <http://nbdpn.org/> [accessed 2016-08-15] [[WebCite Cache ID 6jmiUhiu3](#)]
  30. Williams LJ, Mai CT, Edmonds LD, Shaw GM, Kirby RS, Hobbs CA, et al. Prevalence of spina bifida and anencephaly during the transition to mandatory folic acid fortification in the United States. *Teratology* 2002 Jul;66(1):33-39. [doi: [10.1002/tera.10060](#)] [Medline: [12115778](#)]
  31. Miller L, Romitti P, Cunniff C, Druschel C, Mathews KD, Meaney FJ, et al. The muscular Dystrophy Surveillance Tracking and Research Network (MD STARnet): surveillance methodology. *Birth Defects Res A Clin Mol Teratol* 2006 Nov;76(11):793-797. [doi: [10.1002/bdra.20279](#)] [Medline: [17036307](#)]



32. Centers for Disease Control and Prevention. Centers for Disease Control and Prevention. Epi Info URL: <http://www.cdc.gov/epiinfo/> [accessed 2016-01-11] [WebCite Cache ID 6eSTFMjtg]
33. SAS Institute, Inc. Procedures Guide: Statistical Procedures. Second Edition. In: Statistical Procedures. Second Edition. Cary, NC: SAS Institute Inc; 2013.
34. United States Census Bureau. 2014. QuickFacts URL: <http://www.census.gov/quickfacts/table/PST045215/00.20173> [accessed 2016-01-11] [WebCite Cache ID 6eSTPh7Tr]
35. 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 and Surveillance* 2015;1(2):e15. [doi: [10.2196/publichealth.4991](https://doi.org/10.2196/publichealth.4991)] [Medline: [27227133](https://pubmed.ncbi.nlm.nih.gov/27227133/)]

## Abbreviations

**BDIS:** birth defects information system

**CDC:** Centers for Disease Control and Prevention

**CRCSN:** Colorado Responds to Children with Special Needs

**DHEC:** Department of Health and Environmental Control

**ICD-9:** International Classification of Diseases, Ninth Revision,

**ICD-9-CM:** International Classification of Diseases, Ninth Revision, Clinical Modification

**ICD-10:** International Classification of Diseases, Tenth Revision,

**IRB:** institutional review board

**KFMC:** Kansas Foundation for Medical Care

**MD STARnet:** Muscular Dystrophy Surveillance, Tracking, and Research Network

**RFA:** Revenue and Fiscal Affairs

*Edited by G Eysenbach; submitted 11.01.16; peer-reviewed by J Salemi, A Montserrat; comments to author 03.04.16; revised version received 01.07.16; accepted 20.07.16; published 29.08.16.*

*Please cite as:*

*Reichard A, McDermott S, Ruttenber M, Mann J, Smith MG, Royer J, Valdez R*

*Testing the Feasibility of a Passive and Active Case Ascertainment System for Multiple Rare Conditions Simultaneously: The Experience in Three US States*

*JMIR Public Health Surveill* 2016;2(2):e151

URL: <http://publichealth.jmir.org/2016/2/e151/>

doi: [10.2196/publichealth.5516](https://doi.org/10.2196/publichealth.5516)

PMID: [27574026](https://pubmed.ncbi.nlm.nih.gov/27574026/)

©Amanda Reichard, Suzanne McDermott, Margaret Ruttenber, Joshua Mann, Michael G Smith, Julie Royer, Rodolfo Valdez. Originally published in *JMIR Public Health and Surveillance* (<http://publichealth.jmir.org>), 29.08.2016. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in *JMIR Public Health and Surveillance*, is properly cited. The complete bibliographic information, a link to the original publication on <http://publichealth.jmir.org>, as well as this copyright and license information must be included.

Original Paper

# IBM Watson Analytics: Automating Visualization, Descriptive, and Predictive Statistics

Robert Eugene Hoyt<sup>1\*</sup>, MD, FACP; Dallas Snider<sup>2\*</sup>, PhD; Carla Thompson<sup>3\*</sup>, EdD; Sarita Mantravadi<sup>1\*</sup>, MS, MPH, CPH, CHES, PhD

<sup>1</sup>College of Health, Department of Health Sciences and Administration, University of West Florida, Pensacola, FL, United States

<sup>2</sup>College of Science and Engineering, Department of Computer Science, University of West Florida, Pensacola, FL, United States

<sup>3</sup>College of Education and Professional Studies, Community Outreach Research and Learning (CORAL) Center, University of West Florida, Pensacola, FL, United States

\* all authors contributed equally

**Corresponding Author:**

Robert Eugene Hoyt, MD, FACP

College of Health

Department of Health Sciences and Administration

University of West Florida

11000 University Parkway

Pensacola, FL, 32514

United States

Phone: 1 8503845235

Fax: 1 850 474 2173

Email: [rhoyt@uwf.edu](mailto:rhoyt@uwf.edu)

## Abstract

**Background:** We live in an era of explosive data generation that will continue to grow and involve all industries. One of the results of this explosion is the need for newer and more efficient data analytics procedures. Traditionally, data analytics required a substantial background in statistics and computer science. In 2015, International Business Machines Corporation (IBM) released the IBM Watson Analytics (IBMWA) software that delivered advanced statistical procedures based on the Statistical Package for the Social Sciences (SPSS). The latest entry of Watson Analytics into the field of analytical software products provides users with enhanced functions that are not available in many existing programs. For example, Watson Analytics automatically analyzes datasets, examines data quality, and determines the optimal statistical approach. Users can request exploratory, predictive, and visual analytics. Using natural language processing (NLP), users are able to submit additional questions for analyses in a quick response format. This analytical package is available free to academic institutions (faculty and students) that plan to use the tools for noncommercial purposes.

**Objective:** To report the features of IBMWA and discuss how this software subjectively and objectively compares to other data mining programs.

**Methods:** The salient features of the IBMWA program were examined and compared with other common analytical platforms, using validated health datasets.

**Results:** Using a validated dataset, IBMWA delivered similar predictions compared with several commercial and open source data mining software applications. The visual analytics generated by IBMWA were similar to results from programs such as Microsoft Excel and Tableau Software. In addition, assistance with data preprocessing and data exploration was an inherent component of the IBMWA application. Sensitivity and specificity were not included in the IBMWA predictive analytics results, nor were odds ratios, confidence intervals, or a confusion matrix.

**Conclusions:** IBMWA is a new alternative for data analytics software that automates descriptive, predictive, and visual analytics. This program is very user-friendly but requires data preprocessing, statistical conceptual understanding, and domain expertise.

(*JMIR Public Health Surveill* 2016;2(2):e157) doi:[10.2196/publichealth.5810](https://doi.org/10.2196/publichealth.5810)

**KEYWORDS**

data analysis; data mining; machine learning; statistical data analysis; natural language processing

## Introduction

Studies have shown that physicians tend to lack data analytical expertise, most likely due to insufficient training in statistics while in medical school or not using statistics on a regular basis [1,2]. Gigerenzer et al conclude that there is “collective statistical illiteracy” in health care today [3]. Moreover, there is evidence that teaching statistics to students outside of their health care specialty area has been relatively unsuccessful [4]. Contrastingly, the need for data analytical expertise by health care personnel has grown exponentially, given the explosion of health care data in the past decade [5]. The situation is complicated by a lack of trained data scientists to analyze new data. According to McKinsey Global Institute, there will be a shortage of 140,000 to 180,000 data scientists by 2018 [6].

Therefore, new tools are needed to assist health care workers in analyzing health data. Clinicians and other health care workers would benefit from tools that could produce descriptive, predictive, and visual analytics more rapidly and easily than tools currently available in most analytical software packages and with little training required for users.

One new tool with potential benefit to health care workers is IBM Watson Analytics (IBMWA), introduced in 2015. Unlike the Watson that won *Jeopardy* in 2011, this platform is not based on cognitive computing; instead, it is based on a practitioner approach to using advanced statistics [7]. IBMWA is available in free, personal, and professional versions. An IBMWA Academic Program was also released in 2015 that provides free access to the professional version of the program for faculty and students, if used for noncommercial purposes [8].

In this paper, we will report on the features of IBMWA and discuss how this software subjectively and objectively compares with other data mining programs.

## Methods

### Software Packages

#### *IBM Watson Analytics*

IBMWA utilizes advanced statistics based on Statistical Package for the Social Sciences (IBM SPSS, IBM Corporation), and the statistical tests used are enumerated in Table 1 [7,9]. The available tests were chosen to provide analysis of nonparametric (nominal and ordinal) and parametric (interval and ratio) types of data. Statistical significance is presented with standard *P* values in addition to effect size.

The features available in the professional (and academic) versions are summarized in Table 2. IBMWA also provides access to additional data sources inherent in social media (such as Twitter feeds) and an enhanced ability to share datasets.

New features are added frequently. IBMWA has 4 basic sections: refine, explore, predict, and assemble that are described in the following paragraphs [7].

#### **Refine**

The refine section is used for data exploration and manipulation. This is a logical starting point to examine any dataset. Here, spreadsheet rows and columns are displayed. Attributes can be renamed, calculations can be embedded, and data can be placed in groups or hierarchies for subgroup analysis. Attributes can be organized into ascending or descending order, and a data score and percent missing data per column is displayed.

#### **Explore**

The explore section is used for descriptive analytics and is demonstrated in the next section. Using the natural language processing (NLP) function of Watson Analytics, a user can enter other questions in the search window. In addition to the map view, data can be represented in tree, heat, grid, area, bar, bubble, line, pie, and categorical charts. IBMWA recommends the optimal display type. The page can be saved for a dashboard, or shared via email, social media, or downloaded. In addition, a hyperlink can be created for the page for remote viewing.

#### **Predict**

The predict section is used for predictive analytics. The user selects the target attribute and IBMWA generates a predictive strength. For categorical targets, the predictive strength is the proportion of correct classifications, and for continuous targets it is 1-relative error. Data quality is rated at the top of the page with mention of any outliers, skewed distributions, and missing data. A user can request an analysis with a single factor, 2 factors, or all factors. The predict results can be saved and shared, similar to the explore function. Hyperlinked statistical details are available that provide the statistical test used, the statistical significance, and effect size.

#### **Assemble**

The assemble section allows for data visualization and dashboard creation. This function creates dashboards, infographics and slide shows by simply dragging and dropping data into the active panes. Multiple choices exist for users to have options for representing or displaying data [7].

#### *IBM Statistical Package for the Social Sciences*

SPSS is a comprehensive statistical package available in standard, professional, and premium versions. Table 3 outlines the features of the professional version [9]. The basic or standard SPSS software package provides a user-friendly, menu-driven approach for analyzing data without the need for programming skills or formula-driven guidelines, options for applying multiple analyses, and generating graphs and visuals to meet the needs of social science researchers with minimal statistical preparation skills. Unlike Microsoft Excel, SPSS provides a direct use and understanding of variables and analyses relative to research principles and interpretation options rather than a formula-driven and spreadsheet (location-dependent) approach to data analyses.

**Table 1.** IBM Watson statistical tests.

Statistical test	Indication
Analysis of variance (ANOVA)	ANOVA tests mean differences among 2 or more groups and whether the mean target value varies across combinations of categories of 2 inputs; If the variation is significant, there is an interaction effect
Asymmetry index	Ratio of skewness to the standard error
Chi-square automatic interaction detector classification tree	Decision tree using chi-square for prediction
Chi-square automatic interaction detector regression tree	Decision tree using chi-square and regression for prediction
Chi-square tests	Using chi-square to compare frequencies in groups, independence, and marginal distributions
D'Agostino's K-squared test of normality	Determines if normal distribution is present
Distribution test	Chi-square test compares conditional distributions with overall distribution
Fisher r-to-t test	Transforms Pearson's <i>r</i> to <i>t</i> test for significance
High low analysis	Partitions categories into high or low groups for analysis
Influence test	Chi-square test determines whether the number of records in a group is significantly different from the expected frequency.
Model comparison test	Tests whether the key driver has an effect on the logistic regression
Paired samples <i>t</i> test	Dependent <i>t</i> test checks whether the means of 2 continuous fields are statistically different or if there is a change in means over time for one group
Unusually high or low analysis	Determines which categories or combinations of categories across categorical fields have unusually high or low target mean values

**Table 2.** Features of IBM Watson Analytics Professional.

Features	IBM Watson Analytics Professional
Maximum number of rows per dataset	10,000,000
Maximum number of columns per dataset	500
Input in .csv, .xls or .xlsx formats	Uploaded from PC, Dropbox, IBM Cognos, Box, and Microsoft OneDrive
Data connections	IBM Cognos BI server, IBM dash DB, IBM DB2, IBM SQL, Microsoft SQL server, MySQL, Oracle, and PostgreSQL
Storage	100 GB; can be increased in increments of 50 GB

**Table 3.** Features of SPSS.

Tool	Function
Core stats and graphics	Standard statistical tests for nominal, ordinal, interval, and ratio data
Integration with R and Python languages	Expands programmability involving additional languages
Multiple linear and mixed modeling	Analyze complex relationships
Nonlinear regression	Predictions on nonlinear data
Simulation modeling	Build risk models when inputs are uncertain
Geospatial analytics	Integrate and analyze time and location data
Customized tables	Analyze and report on numerical and categorical data
Charts, graphs, and mapping	Assist reporting capabilities
Missing value analysis	Address missing data, imputation, etc
Advanced data preparation	Identify data anomalies
Decision trees	Identify group relationships to predict future events
Forecasting techniques	Predict trends with time-series data
SPSS Text Analytics	This add-on complementary software package accompanies SPSS to provide qualitative data analyses and visuals for quantitative data simultaneously analyzed with SPSS

**Table 4.** Features of Microsoft Excel Analysis ToolPak.

Tool	Function
Analysis of variance (ANOVA)	Determines variance on single or multiple factors and mean differences among 2 or more groups
Correlation	Determines if a pair of variables are related
Covariance	Determines if a pair of variables move together and mean differences in 2 or more groups when controlling for initial group differences
Descriptive statistics	Determines central tendency and variability in the data
Exponential smoothing	Predicts a value based on prior forecast
<i>F</i> test for 2-sample variances	Performs a 2-sample <i>F</i> test to compare population variances and mean differences relative to variability testing
Fourier analysis	Transforms time-based patterns into cyclical components
Histogram	Calculates frequencies of values in dataset
Moving average	Forecasts values based on prior averages
Random number generation	Fills a range with independent random numbers
Rank and percentile	Creates a table with ordinal and percentile ranks and used with chi-square analyses
Regression	Linear regression based on “least squares” method
Sampling	Creates a sample from a population
<i>t</i> test	Tests for equality of population means, with equal and unequal variances based on 1-group or 2-group datasets
<i>z</i> test	Performs a 1-sample <i>z</i> test for population comparison or a 2-sample <i>z</i> test for means with known variances

**Table 5.** Features of Microsoft SQL Server Analysis Services.

Tool	Function
Multiple data inputs	Use tabular data, spreadsheets, and text files
Data management	Data cleaning; management; and extract, transform, and load
Model testing	Use cross-validation, lift, and scatter charts
Data mining algorithms	Clustering, Naïve Bayes, decision trees, neural networks, regression, and association rules
Scripting language support	Mining objects are programmable

**Table 6.** Features of Waikato Environment for Knowledge Analysis.

Tool	Function
Preprocess	Descriptive statistics and ability to preprocess data; Data from .csv and .arff files, web data, database data, and ability to generate artificial data
Classify	Classify data from Bayes, neural networks, regression, decision trees, production rules, and other algorithms
Cluster	12 clustering algorithms, to include the common simple k-means
Associate	Association rules for pattern recognition in data
Select attributes	Searches for best set of attributes in dataset
Visualize	Visualization of data into graphs, etc

### **Microsoft Excel Analysis ToolPak**

The ToolPak is a spreadsheet add-on that provides the features found in Table 4 [10]. The use of Microsoft Excel Analysis ToolPak requires the user to acquire and install the add-on as well as cognitively account for the differences between the ToolPak spreadsheet-driven program (involving placeholders for data locations requiring user-generated formulas to drive analyses) and SPSS or IBMWA analysis-driven programs, whereby data and locations are defined variables with analyses formulas inherent in the software programs.

### **Microsoft SQL Server Analysis Services**

Microsoft SQL Server Analysis Services is an integrated platform for data mining that uses relational or cube data in multiple formats to provide predictive analytics. A summary of the features is provided in Table 5. Mining choices include clustering, neural networks, decisions trees, and custom plug-in algorithms. Because Analysis Services is a component in SQL Server's suite of business intelligence tools, it integrates easily with the SQL Server database engine and reporting services component. Analysis Services provides a confusion matrix, but requires the user to manually calculate classification accuracy measures, such as precision and recall. Additionally, Analysis Services generates a lift chart, in lieu of receiver operating characteristic curves [11].

### **Waikato Environment for Knowledge Analysis**

Waikato Environment for Knowledge Analysis (WEKA) is a free machine-learning software platform developed by the University of Waikato in New Zealand. This popular program is used for data mining utilizing primarily classification and clustering tools consisting of rules, decision trees, and multiple other algorithms. WEKA calculates true positive rates, false positive rates, precision, and recall. WEKA will also create the receiver operator characteristic curves and area under the curve [12]. A summary of the features in WEKA is provided in Table 6.

### **Data**

The datasets used for comparing IBMWA and prevailing software are publicly available datasets.

### **County Health Rankings for Florida**

The dataset used to demonstrate IBMWA features was derived from the publicly available 2014 County Health Rankings for the state of Florida [13]. The dataset was altered by merging demographics with health factors and deleting confidence intervals and zscores to reduce the number of attributes (columns). This dataset ranks counties in each state by health outcomes and health factors, with health factors categorized as health behavior, clinical care, social and economic factors, and physical environment. The data matrix was comprised of columns consisting of 41 common health factor attributes and demographics; the rows consisted of the 67 Florida counties.

### **Heart Disease Data**

The dataset used for comparison among the analytical software packages was derived from a well-known and validated machine-learning repository [14]. The selected dataset focused on the diagnosis of heart disease (presence or absence) based on 13 common cardiac risk factors or test attributes (columns) and 270 instances or rows (patients). The thallium test (cardiac scan) attribute had 3 subcategories: 3 = normal test, 6 = fixed defect, and 7 = reversible defect.

## **Results**

### **Use Cases**

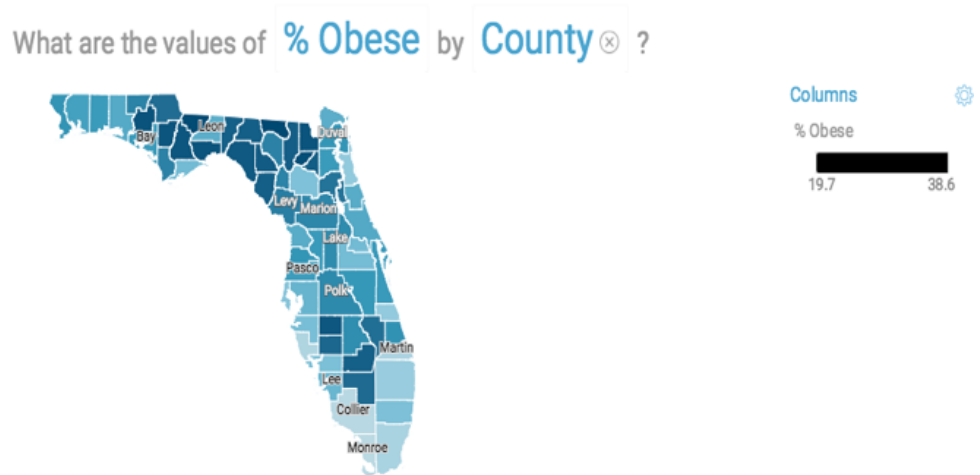
#### **Explore Option**

The use cases shown in this section are generated using the sample data file named 2014 County Health Rankings for the State of Florida. In the explore section, which is used for descriptive analytics, Watson Analytics automatically generated 10 questions based on the data such as "What is the breakdown of % obese by county?" A map of all Florida counties was automatically generated (without user prompting) with % obese noted for each county (Figure 1), as well as the range. The user can mouse over each county for specific data.

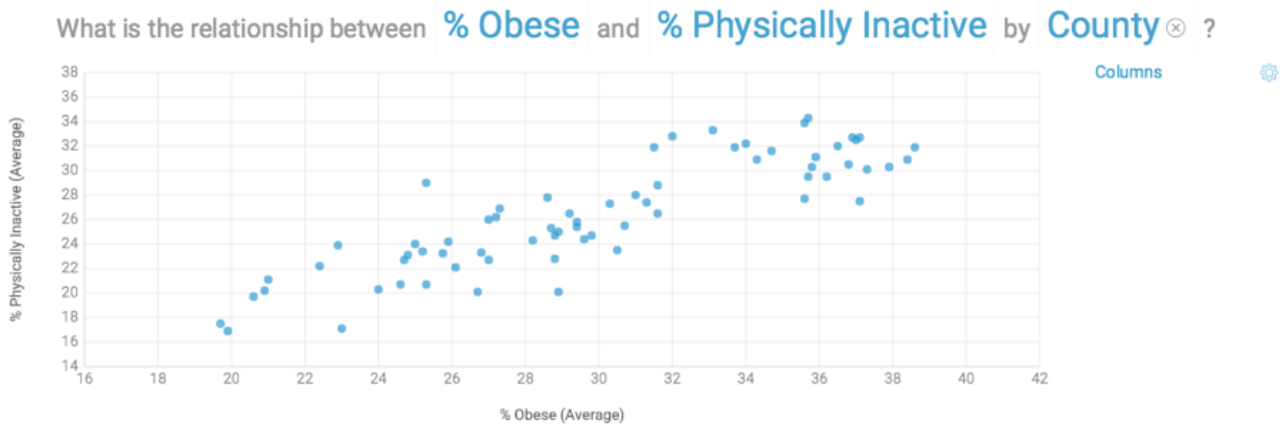
A user can also enter questions in the search window by leveraging the NLP function of Watson Analytics, for example, "What is the relationship between % physically inactive and % obese by county?" (Figure 2)



**Figure 1.** The % obese by Florida county.



**Figure 2.** The relationship between % physically inactive and % obese by county.

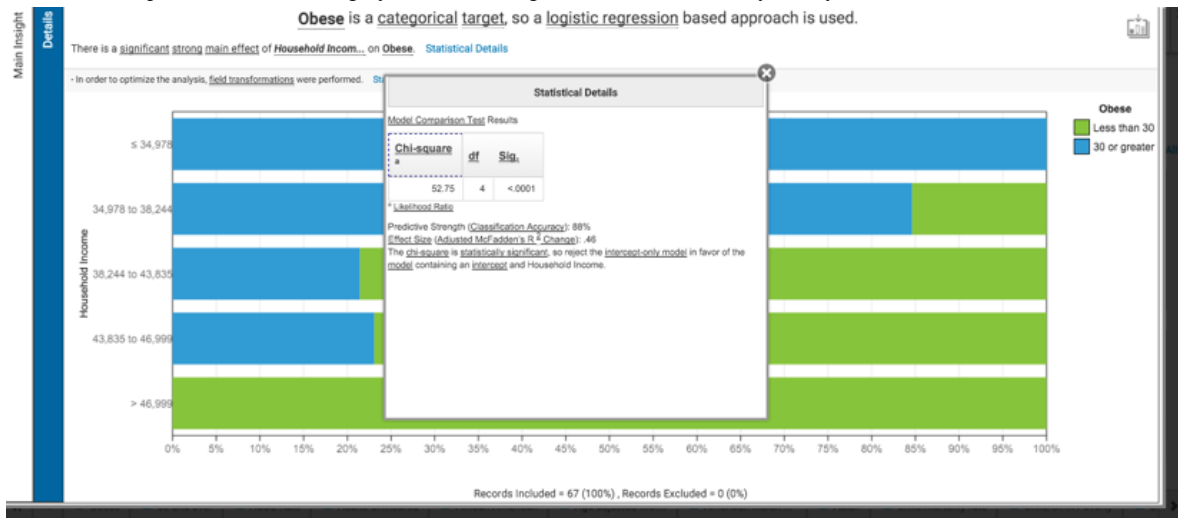


**Figure 3.** Predictors for factors related to % obese.





**Figure 4.** Predict option for nominal category of less than or greater than 30% obese by county.



**Figure 5.** Dashboard of Florida County Health Rankings.



**Predict Option**

The predict option in IBMWA is utilized for predictive analytics. In our analysis of the 2014 County Health Rankings for the state of Florida, 74 associations were noted at the top of the page. The attribute “children in poverty” was associated with “teen birth rate.” Select “statistical details” and a Pearson correlation of .79 with  $P < .001$  and effect size of 0.63 was noted.

When % obese was selected as the target, predictions were automatically generated. The top predictor for “% obesity” was “% physically inactive” at 69%, but IBMWA recommended the addition of “% African-American,” which increased the predictive ability to 85%. A screenshot of the predict function results is shown in Figure 3.

The “% obesity” column of attributes was then subdivided into counties with less than or more than 30% obesity reported and the predict function was reexecuted. This second analysis used logistic regression and produced household income as the strongest predictor at 88% predictive strength. A chi-square analysis comparing the categorical variables demonstrated the following:  $P < .001$ , effect size = .46 (Figure 4).

**Assemble Option**

The assemble option contains functionality to create dashboards, infographics, and slide shows. An example of an IBMWA interactive dashboard display using the dataset, 2014 County Health Rankings for the state of Florida reflecting the “% obesity” by Florida County is depicted in Figure 5. The county dropdown list is active; such that once a county is selected, the percentages change and correspondingly the map changes.

**Comparative Study**

The results from the comparative study among the software packages are presented in the following subsections. The same heart disease dataset was used as the input to each software package and each package provides differing statistics and measures which are summarized in Table 7.

**IBM Watson Analytics**

The IBMWA software conducted a logistic regression for classification purposes. When the target attribute of heart disease (present or absent) was used in IBMWA, it revealed that the thallium test had a predictive strength of 76% (percent correct classification). The thallium test attribute had 3 subcategories: 3 = normal, 6 = fixed defect, and 7 = reversible defect. Based

on either normal exam or reversible defect on thallium testing, the chi-square test revealed  $P < .001$  and an effect size of 0.48 for the normal test and 0.63 for the reversible defect.

When a full model (3 variables) analysis is conducted with logistic regression, the software also conducts a likelihood ratio test (chi-square) to determine if the addition of the variables improves the fit of the model. Predictive strength increases to 80% (percent correct classification) and statistical significance of the target predictor of thallium reduced. Interactions between thallium and the number of vessels calcified on fluoroscopy were not significant,  $P < .09$ . The likelihood ratio test ( $\chi^2_1 = 11.08$ ,  $P < .08$ ) was not statistically significant to the 5% significance level, and thus the reduced model of thallium alone and heart disease is the best fit.

### ***IBM Statistical Package for the Social Sciences***

Binary logistic regression with heart disease as the dependent variable and thallium as a single predictor was conducted. As confirmed in the IBMWA results, predictive strength and percent correctly classified increases as more variables are included in the regression; however, statistical significance reduces.

Logistic regression (LR) with 3 predictors—thallium, number of vessels calcified on fluoroscopy, and the interaction effect—was conducted, illustrating that the predictive strength of the model was 78%, and the interaction effect was not significant. The number of vessels calcified by fluoroscopy and the thallium test variables were statistically significant with  $P = .04$  and  $P < .001$ , respectively. The LR test compared with the intercept only model was significant, with  $\chi^2_3 = 120.5$  and  $P < .001$ , indicating that the 3 variable model improved model fit over the intercept only model.

Thereafter, forward selection using the LR test was also conducted for appropriate variable selection, reducing collinearity and demonstrating model fit. By the end of the stepwise forward regression concerning all variables, the LR test indicated that thallium remained a statistically significant predictor, as well as gender, type of chest pain, electrocardiogram results, exercise-related angina, ST wave depression, and number of vessels calcified by fluoroscopy. Percent correctly classified increased to 90%. The variables gender ( $\chi^2_1 = 3.9$ ,  $P = .049$ ), exercise-related angina ( $\chi^2_1 = 5.7$ ,  $P = .02$ ), and electrocardiogram results were statistically significant at the 5% level, whereas, types of chest pain ( $\chi^2_1 = 13.3$ ,  $P < .001$ ), ST wave depression ( $\chi^2_1 = 11.7$ ,  $P = .001$ ), number of vessels calcified by fluoroscopy ( $\chi^2_1 = 19.9$ ,  $P < .001$ ), and thallium ( $\chi^2_1 = 15.1$ ,  $P < .001$ ) were statistically significant at the 1% level. Comparing this full model with the intercept only model, it was found that  $\chi^2_1 = 77.8$  and  $P < .001$ . These results illustrate that the full model had a better model fit than

both the intercept only model. However, IBMWA uses the LR test to compare models with the reduced fit single predictor model as the default setting, whereas SPSS uses the baseline, intercept only model as the default setting for LR test comparison.

A chi-square analysis was also performed using SPSS with a resulting likelihood ratio of 78% for comparison purposes. Based on the normal exam or reversible defect on thallium testing, the chi-square test revealed a significant relationship ( $\chi^2_1 = 76.1$ ,  $P < .001$ ) and an effect size of 0.53 for the normal test and 0.67 for the reversible defect test.

### ***Microsoft Excel Analysis ToolPak***

The ToolPak software can only conduct linear regression, not logistic regression for classification.

Analysis was not performed because a chi-square test would have to be manually run between the target attribute and each column. The expected values would need to be calculated and run against the actual values to arrive at the chi-square result and  $P$  value. This is very labor intensive compared with the other platforms tested.

### ***Microsoft SQL Server Analysis Services***

Data were analyzed using a decision tree and neural network to compare for classification accuracy. To train the classifier models, 70% of the data was used, whereas the remaining 30% was held out for testing. The decision tree algorithm was chosen because of the ease of understanding the results, while a neural network was selected because of the ability to generally produce better classification results. The decision tree yielded a sensitivity of 0.80 and specificity of 0.78, while neural networks yielded a sensitivity of 0.77 and a specificity of 0.92. Both algorithms have parameters that can be adjusted to improve classification accuracy; however, these parameters need to be adjusted cautiously to avoid “overfitting” the model.

### ***Waikato Environment for Knowledge Analysis***

A J48 decision tree was used as the algorithm with 10-fold cross validation. The outcome was correctly classified 78% of the time. The precision for the presence of heart disease was 0.931 and recall (sensitivity) was 0.628. Precision for the absence of heart disease was 0.692 and the recall was 0.947.

### **Summary**

These preliminary informal analyses indicate that the 4 analytical programs provide similar results using the same dataset. WEKA does provide a confusion matrix, Kappa statistic, and receiver operator characteristics curve area statistic, with neither of these analytics supplied by IBMWA. WEKA, in contrast to IBMWA, includes more than 50 different algorithms, without any recommendations regarding the optimal choice.

**Table 7.** Results of the comparison of different analytical packages.

Software package	Results
IBM Watson Analytics	Using logistic regression, the thallium test had a predictive strength of 76% (percent correct classification); chi-square test revealed $P < .001$ and an effect size of 0.48 (normal test) and 0.63 (reversible defect)
Statistical Package for the Social Sciences	Using logistic regression, a full model with thallium, number of vessels calcified on fluoroscopy, and interaction test increased the predictive strength to 78%; however, a statistically insignificant chi-square test proved that the single model using thallium had the better model fit. $\chi^2_1 = 76.1$ , $P < .001$ ; effect size 0.53 (normal test) and 0.67 (reversible defect)
SQL Server Analysis Services	Decision tree analysis yielded a sensitivity of 0.80 and specificity of 0.78, while neural networks yielded a sensitivity of 0.77 and a specificity of 0.92
Waikato Environment for Knowledge Analysis	Decision tree precision for presence of heart disease was 0.93 and recall (sensitivity) was 0.63; precision for absence of heart disease was 0.69 and recall was 0.95

## Discussion

### Principal Findings

According to Dr. Bill Hersh, “Analytics and related activities are the future of clinical informatics, realizing the goal of my definition of the field, which is the use of information to improve individual health, health care, public health, and biomedical research [15].” To achieve this goal, we will need a well-trained workforce and supportive analytical tools.

IBMWA is an analytical program based on SPSS that automatically generates descriptive, predictive, and visual analytics. This approach is compatible with “greater statistics” proposed by John Chambers in 1993 [16]. Traditional statistics can be laborious when manual computation of complicated formulas is required. Steps to simplify this approach will likely be well received. One could argue that automating analytics is a logical progression, similar to using electronic medical calculators, instead of paper calculations or using the “what you see is what you get” program to create webpages, rather than requiring html programming. This program will likely open the analytics playing field to a larger audience of health care workers.

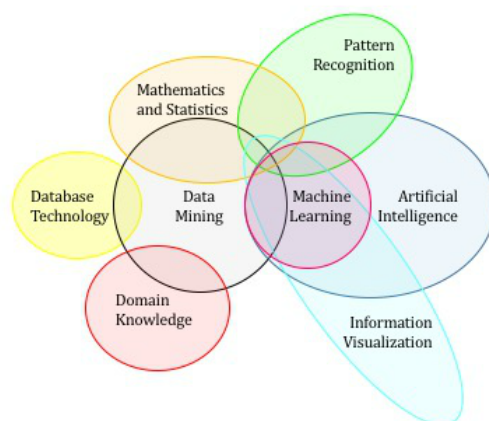
The learning curve for IBMWA is much less steep than Microsoft Excel Analysis ToolPak, SPSS, WEKA, or Microsoft Server Analysis Services. Instead of needing an extensive background in statistics to decide on the statistical method of choice, this is performed automatically for the user. A busy health care worker might use this program to gain preliminary results and then consult an expert in data science or statistics.

IBMWA is able to handle very large datasets and applies the most common statistical tests required, but does not perform

data mining using machine-learning techniques, such as neural or Bayes networks, and is not appropriate for many big data sets. The statistical approach is, however, complimentary to the machine-learning approach. Statistical modeling using a program such as IBMWA usually involves smaller datasets, a hypothesis, and a list of assumptions. Machine-learning, on the other hand, can handle larger datasets, and does not require the same hypotheses or assumptions. An overview of the existing software programs supports IBMWA as belonging in the overlap region between data mining and statistics as demonstrated in [Figure 6](#). In addition, this preview of the use of multiple software programs initiates a scholarly conversation concerning the considerable overlap between statistics and machine learning as exemplified in [Figure 6](#) [17].

A comparison of IBMWA with 3 other data analytical software resulted in similar, but not identical results. We did not report results with Microsoft Excel due its inability to perform logistic regression and the labor intensive nature of the analysis.

IBMWA may be a helpful adjunct approach to teaching both statistics and data mining, given its speed, functionality, and ease of use. Case studies could be presented in the domain of interest and both the clinical results and statistical methods could be discussed. The average user would be able to see missing data, skewed distributions, and outliers with minimal effort. Large datasets are amenable to data analyses and quick response outcomes using IBMWA- a key element in teaching and learning statistical concepts. In the University of West Florida, IBMWA is used to augment understanding and applications of statistical concepts in several health informatics and computer science courses. The use of NLP to help explore, analyze, and visualize datasets would be helpful for most graduate students, regardless of field.

**Figure 6.** Overlap between statistics and machine learning.

### Limitations

For classification purposes with categorical data, IBMWA does not offer the user a choice of statistics, but rather selects the approach for the user (eg, logistic regression). IBMWA did not aid users in interpreting results, especially the impact of unrelated variables and highly correlated variables (multicollinearity). The aim of IBMWA is ease of use but correct interpretations might be eluded due to the lack of odds ratios, confidence intervals, and an explanation for seemingly counterintuitive results.

The combination option in IBMWA makes it easier for users to create complex models, evaluating the effect of up to 3 variables on the dependent variable. Users may not be able to determine the best model; for the average user, it becomes harder to determine if the right model is chosen—with the appropriate theoretically related variables. For example, with multiple predictors in a classification analysis, the chi-square value provided indicates if model fit improves with complexity and increased variables (likelihood ratio test result). Predictive strength increases as additional variables are added; however, this is counterintuitive and easily confused as decreased statistical significance (low chi-square value) with increased predictors—leading to inaccurate interpretation. In IBMWA, as variables are added and complexity increases, the model starts to become confusing to interpret without taking the extra step of checking the fine print definitions; the chi-square and *P*-value provided in a full model is actually for comparing goodness of fit of a statistical model (LR test), not a measure of variable statistical significance. Users may initially focus on the fact that predictive strength will increase each time another variable is added to the model, resulting in a model that is overly complex with unrelated variables. For early health data analysts, the issue of highly correlated variables (multicollinearity) and model fit is hard to detect and interpret. Too many variables can result in predicting noise, rather than the dependent variable (overfitting). The LR test acts as a warning sign for users—don't add more variables in the model just to increase predictive strength.

In addition, IBMWA provides for ease of data exploration, creating a conundrum for the average user—with such ease of data mining and exploration, eventually the user will detect chance correlations between variables that appear to be significant relationships. IBMWA allows the average user to depend on statistical significance as a measure of the strength of the correlation, as well as the correct model; however, a correctly specified model might have insignificant, theoretically relevant predictors. The IBMWA software does not provide statistical significance for each predictor but only the results of the model comparison test (likelihood ratio test).

Limitations exist with every analytical platform. Data preprocessing (imputation and data quality) is needed and is critical for success. It is estimated that about 80% of the time spent analyzing data, is spent exploring and preparing the data for analysis [18]. Specifically, data may need cleaning, integration, reduction, or transformation before analysis procedures [19]. In addition, data must be of sufficient quality and volume to answer a question. Finally, domain expertise and appropriately prepared researchers are needed to ask appropriate questions and interpret results, regardless of the analytical platform used.

### Conclusions

IBMWA is a new and interesting analytical tool that may be of value to multiple types of health care workers; however, no statistical program will replace the time needed for preprocessing data and asking pertinent questions regarding the dataset, but the time spent on analytical processes will be greatly expedited. The IBMWA approach needs to be compared and contrasted with other approaches and by a diverse group of users to better understand its role within the analytics realm. Clearly, IBMWA has limitations but IBM is making frequent changes to this program so users can expect more functionality in the future. Additionally, IBMWA may motivate educators and practitioners to question if it is potentially effective as an adjunct in teaching statistics and analytics in health care.



## Conflicts of Interest

None declared.

## References

1. Wegwarth O, Schwartz LM, Woloshin S, Gaissmaier W, Gigerenzer G. Do physicians understand cancer screening statistics? A national survey of primary care physicians in the United States. *Ann Intern Med* 2012 Mar 6;156(5):340-349 [FREE Full text] [doi: [10.7326/0003-4819-156-5-201203060-00005](https://doi.org/10.7326/0003-4819-156-5-201203060-00005)] [Medline: [22393129](https://pubmed.ncbi.nlm.nih.gov/22393129/)]
2. Manrai AK, Bhatia G, Strymish J, Kohane IS, Jain SH. Medicine's uncomfortable relationship with math: calculating positive predictive value. *JAMA Intern Med* 2014 Jun;174(6):991-993 [FREE Full text] [doi: [10.1001/jamainternmed.2014.1059](https://doi.org/10.1001/jamainternmed.2014.1059)] [Medline: [24756486](https://pubmed.ncbi.nlm.nih.gov/24756486/)]
3. Gigerenzer G, Gaissmaier W, Kurz-Milcke E, Schwartz LM, Woloshin S. Helping doctors and patients make sense of health statistics. *Psychol Sci Public Interest* 2007 Nov;8(2):53-96 [FREE Full text] [doi: [10.1111/j.1539-6053.2008.00033.x](https://doi.org/10.1111/j.1539-6053.2008.00033.x)] [Medline: [26161749](https://pubmed.ncbi.nlm.nih.gov/26161749/)]
4. Yimaz M. *Amstat*. 1996. The Challenge of Teaching Statistics to Non-Specialists URL: <http://www.amstat.org/publications/jse/v4n1/yilmaz.html> [accessed 2016-01-20] [WebCite Cache ID 6f3StcC1m]
5. DeGaspari J. Managing the Data Explosion. *Healthcare-informatics*. 2013. URL: <http://www.healthcare-informatics.com/article/managing-data-explosion> [accessed 2016-02-04] [WebCite Cache ID 6f3UHVtLq]
6. Davis P. McKinsey Report Highlights the Impending Data Scientist Shortage Pivotal Blog. 2013 Jul 23. URL: <https://blog.pivotal.io/data-science-pivotal/news/mckinsey-report-highlights-the-impending-data-scientist-shortage> [accessed 2016-02-04] [WebCite Cache ID 6f3TQiGxE]
7. IBM. IBM Watson Analytics URL: <http://www.ibm.com/analytics/watson-analytics/> [accessed 2016-05-20] [WebCite Cache ID 6f3URVTtm]
8. IBM. IBM Watson Analytics Academic Program-WAP URL: <https://watson.analytics.ibmcloud.com/solutions/industry/education/wap> [accessed 2016-06-15] [WebCite Cache ID 6f3V7K6s2]
9. IBM. IBM SPSS Software URL: <http://www-01.ibm.com/software/analytics/spss> [accessed 2016-01-02] [WebCite Cache ID 6f3XEEkOU]
10. Use the Analytics ToolPak to Perform Complex Data Analysis. Microsoft; 2016. URL: <https://support.office.com/en-us/article/Use-the-Analysis-ToolPak-to-perform-complex-data-analysis-f77cbd44-fdce-4c4e-872b-898f4c90c007> [WebCite Cache ID 6f3X2Rz37]
11. Microsoft. Data Mining (SSAS) URL: <https://msdn.microsoft.com/en-us/library/bb510516.aspx> [accessed 2016-02-04] [WebCite Cache ID 6f3XbsRvU]
12. Waikato. WEKA 3: Data Mining Software in Java URL: <http://www.cs.waikato.ac.nz/ml/weka> [accessed 2016-01-02] [WebCite Cache ID 6f3XL1Psy]
13. County Health Rankings. County Health Rankings & Roadmaps: Building a Culture of Health, County by County URL: <http://www.countyhealthrankings.org/rankings/data/fl> [accessed 2015-05-20] [WebCite Cache ID 6f3VUACLq]
14. UCI. Heart Disease Data Set URL: <https://archive.ics.uci.edu/ml/datasets/Heart+Disease> [accessed 2016-01-02] [WebCite Cache ID 6f3WAvzRv]
15. Hersh W. A stimulus to define informatics and health information technology. *BMC Med Inform Decis Mak* 2009 May 15;9:24 [FREE Full text] [doi: [10.1186/1472-6947-9-24](https://doi.org/10.1186/1472-6947-9-24)] [Medline: [19445665](https://pubmed.ncbi.nlm.nih.gov/19445665/)]
16. Chambers JM. Greater or lesser statistics: a choice for future research. *Stat Comput* 1993 Dec;3(4):182-184. [doi: [10.1007/BF00141776](https://doi.org/10.1007/BF00141776)]
17. Witten I, Frank E, Hall M. Introduction to data mining. In: *Data Mining: Practical Machine Learning Tools and Techniques*. 3rd edition. Burlington, MA: Morgan Kaufmann-Elsevier; 2011.
18. Donoho D. 50 Years of Data Science. *Csail.mit*. 2005 Sep 18. URL: <http://courses.csail.mit.edu/18.337/2015/docs/50YearsDataScience.pdf> [WebCite Cache ID queryphp]
19. Jiawei H, Micheline K, Jian P. Data preprocessing. In: *Data Mining: Concepts and Techniques*. 3rd edition. Waltham, MA: Morgan Kaufmann-Elsevier; 2012.

## Abbreviations

**ANOVA:** Analysis of variance  
**IBM:** International Business Machines Corporation  
**IBMWA:** IBM Watson Analytics  
**LR:** Logistic regression  
**NLP:** natural language processing  
**SPSS:** Statistical Package for the Social Sciences  
**WEKA:** Waikato Environment for Knowledge Analysis



*Edited by G Eysenbach; submitted 29.03.16; peer-reviewed by S Bhattacharya, CH Li, C McGregor, A Ramachandran; comments to author 25.07.16; revised version received 31.08.16; accepted 17.09.16; published 11.10.16.*

*Please cite as:*

*Hoyt RE, Snider D, Thompson C, Mantravadi S*

*IBM Watson Analytics: Automating Visualization, Descriptive, and Predictive Statistics*

*JMIR Public Health Surveill 2016;2(2):e157*

*URL: <http://publichealth.jmir.org/2016/2/e157/>*

*doi: [10.2196/publichealth.5810](https://doi.org/10.2196/publichealth.5810)*

*PMID: [27729304](https://pubmed.ncbi.nlm.nih.gov/27729304/)*

©Robert Eugene Hoyt, Dallas Snider, Carla Thompson, Sarita Mantravadi. Originally published in JMIR Public Health and Surveillance (<http://publichealth.jmir.org>), 11.10.2016. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Public Health and Surveillance, is properly cited. The complete bibliographic information, a link to the original publication on <http://publichealth.jmir.org>, as well as this copyright and license information must be included.

Original Paper

# Efficacy of Web-Based Collection of Strength-Based Testimonials for Text Message Extension of Youth Suicide Prevention Program: Randomized Controlled Experiment

Phyo Thiha<sup>1</sup>, BA, MS; Anthony R Pisani<sup>2,3</sup>, AB, MEd, PhD (Clinical Psych); Kunali Gurditta<sup>2</sup>, BA (Psych) Hon; Erin Cherry<sup>4</sup>, BS, PhD; Derick R Peterson<sup>5</sup>, BA, MA, PhD; Henry Kautz<sup>1</sup>, AB, MA, MS, PhD; Peter A Wyman<sup>2</sup>, BA, MA, PhD

<sup>1</sup>Department of Computer Science, University of Rochester, Rochester, NY, United States

<sup>2</sup>Department of Psychiatry, University of Rochester School of Medicine and Dentistry, Rochester, NY, United States

<sup>3</sup>Department of Pediatrics, University of Rochester School of Medicine and Dentistry, Rochester, NY, United States

<sup>4</sup>Northrop Grumman, Huntsville, AL, United States

<sup>5</sup>Department of Biostatistics and Computational Biology, University of Rochester, Rochester, NY, United States

**Corresponding Author:**

Anthony R Pisani, AB, MEd, PhD (Clinical Psych)  
Department of Psychiatry  
University of Rochester School of Medicine and Dentistry  
300 Crittenden Blvd.  
Box PSYC  
Rochester, NY, 14642  
United States  
Phone: 1 585 275 3644  
Fax: 1 585 276 2065  
Email: [anthony\\_pisani@urmc.rochester.edu](mailto:anthony_pisani@urmc.rochester.edu)

## Abstract

**Background:** Equipping members of a target population to deliver effective public health messaging to peers is an established approach in health promotion. The Sources of Strength program has demonstrated the promise of this approach for “upstream” youth suicide prevention. Text messaging is a well-established medium for promoting behavior change and is the dominant communication medium for youth. In order for peer ‘opinion leader’ programs like Sources of Strength to use scalable, wide-reaching media such as text messaging to spread peer-to-peer messages, they need techniques for assisting peer opinion leaders in creating effective testimonials to engage peers and match program goals. We developed a Web interface, called Stories of Personal Resilience in Managing Emotions (StoryPRIME), which helps peer opinion leaders write effective, short-form messages that can be delivered to the target population in youth suicide prevention program like Sources of Strength.

**Objective:** To determine the efficacy of StoryPRIME, a Web-based interface for remotely eliciting high school peer leaders, and helping them produce high-quality, personal testimonials for use in a text messaging extension of an evidence-based, peer-led suicide prevention program.

**Methods:** In a double-blind randomized controlled experiment, 36 high school students wrote testimonials with or without eliciting from the StoryPRIME interface. The interface was created in the context of Sources of Strength—an evidence-based youth suicide prevention program—and 24 ninth graders rated these testimonials on relatability, usefulness/relevance, intrigue, and likability.

**Results:** Testimonials written with the StoryPRIME interface were rated as more relatable, useful/relevant, intriguing, and likable than testimonials written without StoryPRIME,  $P=.054$ .

**Conclusions:** StoryPRIME is a promising way to elicit high-quality, personal testimonials from youth for prevention programs that draw on members of a target population to spread public health messages.

(*JMIR Public Health Surveill* 2016;2(2):e164) doi:[10.2196/publichealth.6207](https://doi.org/10.2196/publichealth.6207)

**KEYWORDS**

mental health; adolescent health; user interface design; suicide; text messaging

## *Introduction*

### **Background and Motivation**

Public health practitioners and researchers are increasingly harnessing mobile technology to enhance the reach and effectiveness of evidence-based interventions [1-3]. Equipping members of a target population to deliver effective public health messaging to peers is an established approach in substance abuse prevention, human immunodeficiency virus prevention, and health promotion [4-7], and is promising for suicide prevention [8]. The effectiveness of peer-to-peer public health messaging strategies is congruent with peer-to-peer influences on behavior [10]. For youth with acute distress and with numerous risk factors for suicide, social modeling of emotional regulation strategies could slow or reverse a negative trajectory or prevent impulsive attempts to escape painful emotions [11,12]. A crucial barrier to technological extensions of these so-called ‘opinion leader’ interventions is the challenge of helping peer leaders to deliver effective messages to peers remotely. This requires development of Web-based replacements for typical in-person components [9]. In this study, we tested a Web-based interface for remotely helping high school peer leaders to produce high-quality, personal testimonials for use in a text messaging extension of an evidence-based, peer-led suicide prevention program.

Sources of Strength is a school-based suicide prevention program that is certified on the United States National Registry of Evidence-Based Programs and Practices. The program seeks to reduce risk and strengthen protective factors in the population by preparing diverse ‘key opinion leaders’ to conduct public health messaging and activities with peers to increase school-wide positive coping norms and communication with trusted adults, and bring suicidal peers to seek for adult help [8,13]. Currently implemented in over 3000 high schools and colleges across the United States, Sources of Strength was the first suicide prevention program to apply a peer ‘opinion leader’ approach. As part of this strategy, peer leaders publicly share personal stories (or testimonials) of “hope, help, and strength” with peers and encourage them to do the same. Those who share their personal stories act as distributed change agents, generating much of the intervention content received by their school’s population. In a randomized controlled trial conducted in 18 schools, trained peer leaders were 4 times more likely to refer a suicidal friend to an adult, and school-wide help-seeking norms improved in schools implementing Sources of Strength [8]. An ongoing trial with 40 schools in rural and underserved communities is testing the impact and proposed network diffusion model of Sources of Strength on reducing suicide attempts.

Text4Strength is a text messaging extension of Sources of Strength. It is currently under development with funding from the National Institute of Mental Health (K23MH101449). Text4Strength uses text-based peer testimonials and other types of text messages to introduce high school students to Sources

of Strength concepts and extends Sources of Strength by also introducing emotional skills associated with decreased depressive symptoms and lower suicide risk [14-17]. Text messaging is a logical medium for disseminating messages because it is the dominant mode of communication for youth [18] and has strong empirical support for delivering health interventions [2,3,19-25]. The current version of Text4Strength specifically targets ninth graders because research indicates an increase in emotional and behavioral problems between ages 14 and 15, when most students are starting high school [26]. In addition to normal developmental changes resulting in heightened emotional activation [27-29] and diminished inhibitory control abilities [30], ninth graders face new challenges associated with the transition to high school, including greater responsibility for seeking help, greater emphasis on romantic relationships, and greater academic pressure.

In developing Text4Strength, we faced challenges that previous work in translating behavioral interventions into text (e.g., [9]) had not yet surmounted: collecting interesting, personal, and credible [31,32] testimonials consistent with the intervention model, goals, and targets. We addressed these challenges by developing Stories of Personal Resilience in Managing Emotions (StoryPRIME), a Web-based interface to guide peer leaders through the process of remembering, sharing, and condensing relevant personal stories into the pithy testimonials needed for text messaging. We developed StoryPRIME through an interdisciplinary design process in which suicide prevention researchers consulted with human-computer interaction (HCI) researchers. Together, we conducted design workgroups with high school peer leaders, as well as feasibility testing with an adult sample to ensure the safety of the StoryPRIME interface prior to testing it in a randomized controlled experiment with a teenage population. In this paper, we report on results from the randomized controlled experiment of StoryPRIME to discover whether or not Web-based eliciting could assist peer leaders in generating high-quality text message testimonials that are interesting, personal, and credible to ninth graders.

### **Development and Feasibility Testing**

Suicide prevention and HCI researchers worked together to translate practices used in Sources of Strength to elicit peer leaders to tell their stories using a Web-based interface. In the Sources of Strength training program, student peer leaders participate in a 4-hour training in which helpful testimonials are modeled and practiced. After initial training, students participate in approximately biweekly meetings where adult advisors give them opportunities to practice and receive feedback sharing their “hope, help, and strength” stories with peers and encourage them to do the same [13]. The challenge we faced was translating this highly organic discussion and coaching process in Sources of Strength into a structured, Web-based interface.

To tackle this challenge, HCI researchers piloted an initial interface with our interdisciplinary team and other lab members,

and then conducted a participatory design [33] session and workgroup with 7 high school peer leaders (4 females, 3 males). We met with the students 7 times, for 60-90 minutes each, to get their feedback on the design of the StoryPRIME interface for testimonial writing and in general, about our plans for the Text4Strength text messaging program. Using this feedback, HCI researchers developed several variations/prototypes of the StoryPRIME interface. We then asked the workgroup participants to write stories about themselves using these interfaces, and discuss the writing experience with the workgroup participants and our research group. Based on the workgroup's feedback, we refined the StoryPRIME interface, partitioning the writing process into 3 main steps to help writers focus on the essential parts of the story: (1) challenge faced, (2) how they solved it, and (3) the outcome.

We pilot tested the StoryPRIME interface with an adult convenience sample in order to ensure feasibility and safety before testing with adolescent high school students. Fifty adults (29 females, 21 males; mean age of 30.28 (standard deviation=7.73)) were recruited from Mechanical Turk (MTurk)—a crowdsourcing website where people are paid to do Web-based tasks. Participants wrote testimonials about common stressful situations and emotional skills and resources they used to cope. A separate sample of 61 adults recruited from MTurk rated the testimonials on how interesting, personal, genuine/credible, and relatable they were. In addition, 2 public health experts rated the stories on how appropriate and (potentially) effective they were. The results were promising: 70% (35/50) of the testimonials were judged as interesting, 78% (39/50) as personal, 80% (40/50) as credible, and 76% (38/50) as relatable. The two suicide prevention experts judged 76% (38/50) and 72% (36/50) of the testimonials as appropriate/safe and effective, respectively. Furthermore, qualitative responses from Web-based participants indicated that they found the experience rewarding and enjoyable, with no negative report of burden or emotional discomfort.

## Methods

### Participants

The participants included, testimonial writers: 36 students (23 females, 13 males) of grades 10-12 (ages 16-18) from 2 rural high schools in Western New York wrote testimonials. All trained Sources of Strength peer leaders in each school (approximately 45) were invited to participate and given parent information letters and permission forms to return; and raters: 24 (10 females, 14 males) ninth grade (ages 14-15) students whose parents gave them permission to participate in this study rated on the testimonials. Two classes of ninth grade students in each school (approximately 50) were invited via to participate via a school-wide announcement and given parent information letters and permission forms to return. All participants were paid US\$10 as a thank you for their time. The University of Rochester institutional review board approved this study.

### Measures

Students rated testimonials on the following dimensions using a 1-5 ('Strongly Disagree' to 'Strongly Agree') Likert scale: relatable ("My friends and I handle similar challenges"),

useful/relevant ("This student's solution/advice would work for me"), intriguing ("I am curious about more detail/background to this SPECIFIC experience"), and likable ("I would be interested in hearing more student experiences like this").

### Design

We tested the StoryPRIME interface using a double-blind, randomized controlled design. Writers were randomly assigned to write a personal story with the aid of StoryPRIME (N=18 with 12 females) or to a control condition (N=18 with 13 females). Within each school, we assigned an equal number of students into either StoryPRIME or control condition—7 per condition for one, and 11 per condition for another school. There was no opportunity for "contamination" because both groups worked on testimonials at the same time in a quiet room on separate computers—participants simply received different on-screen instructions and interface depending upon their condition. We asked student writers to write as many testimonial as they could during a 30-minute session. Most students (N=34) completed writing 2 testimonials while 2 of them wrote just one each in the allotted time. Testimonial writers collectively produced 70 testimonials, 35 in the StoryPRIME condition and 35 in the control condition. At the end of the writing session, we asked students to complete a brief Web-based anonymous survey for their feedback on their writing experience and StoryPRIME interface.

Testimonials were presented in identical format and random order to the ninth grade student raters (N=24 with 10 females) who were blind to the writer's condition. Each testimonial was rated by at least 3 different students. Testimonial raters rated a random selection of 12 testimonials each, 6 in each condition. Ratings were conducted anonymously on private laptops of the raters and they received payment immediately after submitting their responses. In addition to explaining to the raters that we would not know who provided which ratings, instructions emphasized that we desired and needed their honest opinions—positive and negative.

### Conditions

We met with student writers in their school's computer lab where each student was assigned a computer. All student testimonial writers accessed a website developed for this study on which they were invited to write strength-based testimonials that would help and appeal to ninth grade students:

*The Center for the Study and Prevention of Suicide works with high schools to teach teenagers emotional skills so that they stay strong when they face hard times. To do that, we need real examples from people who have used different strategies for healthy responses to conflict or common challenges they face in high school.*

Students then selected a challenge to write about from a list of common challenges faced by high school students, generated with the help of the student workgroup in the participatory design process (see [Textbox 1](#)).

### StoryPRIME condition

Participants in the StoryPRIME condition were asked to:



Think of a time in high school when you handled or learned something about [the challenge user selected previously].

After proceeding, the participants saw:

*Thank you! Now we will ask you THREE questions to help you write the details about the challenge (stress) you just selected. Please be as specific as possible. At the end, we will ask you to summarize your answers into a ~300-character story/testimonial.*

If the participants chose the “Ready to Help” option, they were presented with a series of 3 writing prompts and accompanying tips. Figure 1 shows these writing prompts for the challenge of “Figuring out one’s ‘place’ or fitting in.” After completing all 3 steps, participants were shown all of their previous responses on a single page and asked to generate a concise summary (~300 characters) as shown in Figure 2. Following completion, participants were asked to repeat the procedure for another challenge.

**Textbox 1.** Common challenges (stressors) presented as options to testimonial writers, collected from the student participatory design workgroup.

1. Learning how to balance time
2. Different class schedule from friends
3. School pressure (classes, grades, tests)
4. Not sure how to get help from teachers/adults
5. Getting cut or benched in a sport or favorite activity
6. Figuring out one’s “place” or fitting in
7. Friend or boyfriend/girlfriend drama
8. Family/parent issues
9. Feeling confused, strong feelings, ups and downs

**Figure 1.** The three sequential writing prompts and corresponding clues shown to testimonial writers in the StoryPRIME condition who chose “Figuring out one’s ‘place’ or fitting in” as their challenge.

The figure displays three sequential writing prompts, each with a teal header and a white body containing the prompt and a clue. The prompts are:

- Describe what happened in this situation related to...**  
Figuring your 'place' or fitting in  
Write about the challenging situation. For example, provide full context such as where, what, when, why/how it happened. You will write about how you handled it in the next step.
- Tell us HOW you successfully handled the situation related to...**  
Figuring your 'place' or fitting in  
What did you do to resolve the situation? Or how did you handle it? Did you use any of these resources to solve it: family support, positive friends, mentors, sports, charitable activities, spirituality, medical and mental health professionals (or other resources)? Next, you'll write about the outcome of this solution.
- Share a positive change because of the way you handled the situation regarding...**  
Figuring your 'place' or fitting in  
How did everything go? Did the situation improved? Did YOU change as a result of this? Did you feel good about the outcome? Next, we'll ask you to summarize what you've written.



**Figure 2.** The final step of StoryPRIME in which testimonial writers are shown their responses to the earlier three writing prompts on a single page.

Here is what you've told us so far:

1. You said this about the situation:

It was the first week of ninth-grade and I realized that some of days I didn't have any classes with my friends.

2. You said this about how your handled the situation.

I was able to talk with other people in the classes that I had and realized that I didn't need my friends in all of my classes. I was able to make new friends in the classes that I did have.

3. You said this about how things changed or got better:

Things ended up okay. I was able to change the situation and make it so I did have classes with my new friends. I changed because I became friends with people who I thought I didn't have anything in common with it turned out that I did.

Please summarize what you've written into a brief story (~300-characters) that might help another high school student. It's OK to use abbreviations as you would in a txt. :)

Try to include as much detail as you can in the summary. Most importantly, make sure to include the description of the situation, how you handled it and the outcome in the summary

Characters written: 0

### Control condition

Participants in the control condition were given simple instructions: "In the space below, please share a brief story (~300 characters) about [the challenge user selected previously], which might help another high school student. It's OK to use abbreviations as you would in a txt."

### Statistical Analyses

Linear mixed-effects models were used to model the average of the 4 questions as a function of intervention (StoryPRIME vs control), controlling for rater via random effects. Similar models were used to model the Likert score for each individual question. Restricted maximum likelihood estimation via `lme()` in R  $\times$ 64-bit version 3.1.1 was used to fit all models. We decided a priori in consultation with our biostatistician (DP) to use significance level of .10 in order to match the scientific needs of testing a novel procedure—that is, to mitigate the risk of throwing out a useful procedure based on an overly rigorous (and arbitrary) *P* level [34,35].

## Results

### Principal Findings

Ninth grade students ( $N=24$ ), who were blind to the conditions, found testimonials written with the aid of StoryPRIME ( $N=35$ ) more favorable than those written with the control interface ( $N=35$ ). The mean of the 4 ratings—each using the 1-5 Likert scale where 5 is the maximum favorable score achievable—for testimonials written in the control condition was 3.47 (standard error (SE)=.10), and the mean for testimonials written in the StoryPRIME condition was 3.60 (SE=.10). The difference in mean ratings across conditions was .13 (SE=.08),  $P=.054$ . When questions were examined individually, student ratings on all 4 questions were in the same direction favoring the StoryPRIME condition over the control. As presented in Table 1, ninth grade students, who were blind to the conditions, found testimonials that had been written with the aid of StoryPRIME more relatable, useful/relevant, and likable than those written with the control interface.

**Table 1.** Summary of ratings for testimonials written in the control versus StoryPRIME conditions.

	Control mean (SE <sup>a</sup> )	StoryPRIME mean (SE)	Intervention effect mean (SE)	One-sided <i>P</i> value
Relatability	3.62 (.11)	3.76 (.11)	.14 (.11)	.097
Usefulness/relevance	3.48 (.10)	3.65 (.09)	.17 (.11)	.055
Intrigue	3.33 (.13)	3.42 (.13)	.09 (.09)	.165
Likability	3.44 (.13)	3.58 (.13)	.13 (.10)	.088
Mean of 4 ratings	3.47 (.10)	3.60 (.10)	.13 (.08)	.054

<sup>a</sup>SE: standard error.

## Student Testimonials

The challenges students wrote about centered mostly on academic pressures (grades, tests), time management, and friends. Some were about fitting in and boy/girlfriend drama. Specifically, the breakdown of peer leader testimonials was as follows: 25 testimonials related to school pressure (12 control, 13 intervention); 19 about balancing time (11 control, 8

intervention); 8 regarding different class schedules from friends (4 control, 4 intervention); 3 about feeling confused (3 control); 4 about relationship drama (2 control, 2 intervention); 6 about fitting in (2 control, 4 intervention); 1 about getting cut in sports (1 control); 3 about not knowing how to get help from teachers and adults (3 intervention); and 1 about family/parent issues (1 intervention). Some of the testimonials written in StoryPRIME can be seen in [Textbox 2](#).

**Textbox 2.** Illustrative examples of testimonials generated by StoryPRIME.

*Coming into high school my grades slipped and I knew I needed help but I was too scared to ask for help. After failing to keep my grades up I had to break out my shell and ask my older peers because I knew they knew how to keep their grades up and improve my time management. Their advice helped me. (Challenge (stressor): school pressure)*

*At the start of 9<sup>th</sup> grade I had difficulty handling in homework and sometimes projects. I attempted to do all my school work all at one time and it would leave me mentally exhausted. When my college success class worked on managing time, I started doing better in all my classes. Now I feel good about myself. (Challenge: learning how to balance time)*

*I didn't have many classes with my friends one year but we made it work. I found them in the cafeteria and we still hung out outside of school. If you don't have friends in your classes then it's a chance to make new ones. I made a bunch of new friends and learned to talk to new people. (Challenge: different class schedule from friends)*

*After 8<sup>th</sup> grade, my two best friends had moved and prior to that I didn't talk to many other people. In 9th grade I began talking to people that I normally wouldn't have, and began involving myself in activities so I had more opportunities to meet more people. I ended up making a lot of new friends. (Challenge: fitting in)*

*My grandmother died of old Age on New Years day, at 2 am. I cried non-stop when I found out and eventually after a month of focusing on school activities and support from my my mom, I got over it and accepted it as life. (Challenge: family issue)*

The StoryPRIME interface prompts writers to describe how they applied some common coping strategies—such as family support, mentors, and positive friends as seen in the second prompt of [Figure 1](#). Students wrote stories that mentioned relying on family support (1 in control, 2 in intervention); positive friends (5 control, 11 intervention); mentors (4 control, 14 intervention); sports (2 control, 0 intervention); and none (23 control, 8 intervention).

Most students in both conditions were happy to have shared their stories. For example, when asked:

*How did you feel? Did you find anything upsetting or uncomfortable while writing the story?*

One student in the control condition wrote:

*It was an aha moment for me and realization.*

Another student in the StoryPRIME condition wrote:

*I'm happy that I was able to share my story, in order to provide advice for students who are entering high school and preparing for challenges, as well.*

## Discussion

### Challenges, Achievements, and Feedback

The randomized controlled experiment produced promising results: (1) high school students assisted by StoryPRIME wrote numerous testimonials about diverse challenges such as school pressure (grades, tests), time management, friends, fitting in, and boy/girlfriend drama, (2) compared with the control condition, students assisted by StoryPRIME wrote testimonials that ninth graders found more relatable, useful/relevant, and likable, and (3) students had a positive experience writing testimonials using the StoryPRIME interface based on the overwhelmingly positive feedback they provided in the exit survey. Although the differences between the StoryPRIME testimonials and the control testimonials are small, they are meaningful—especially given the scale at which we could use StoryPRIME to reach potential student testimonial contributors.

Regarding the nonsignificant results with respect to whether messages were intriguing (“I am curious about more detail/background to this SPECIFIC experience”), we suspect

that the suggested 300-character limit might not be sufficient enough to write intriguing (interesting) stories. Allowing students to write more characters may encourage them to include more details about how the event transpired, thereby making the story more intriguing. This goes in line with some suggestions from writers of both StoryPRIME and control condition such as:

*Ask for more details*

*Add more pictures and fun questions such as how are you doing? do you feel okay today? etc.*

*...maybe just letting kids describe their situation more and how family was involved.*

The downside of allowing more details is that the resulting stories may no longer fit the criteria of a short text message, which is what we originally set out to achieve for our text message campaign.

Feedback on StoryPRIME interface by student writers were especially encouraging; one of the writers in StoryPRIME condition wrote:

*I thought this website (StoryPRIME) was great. I liked the fact that it helped me remember some times I struggled by giving me examples of things other students struggled with.*

On the other hand, a writer in control condition suggested:

*more room to write about something or more prompting questions for personalization.*

Also, another wrote:

*It's a bit hard to recall what happened over a years ago and decide which one to write about.*

These comments and better ratings achieved by testimonials written using StoryPRIME suggest that the interface is helping students recall relevant topics in high school and write effective stories about them.

Despite the aforementioned positive feedback, the StoryPRIME interface is far from perfect. A few writers in StoryPRIME condition suggested potential improvements to the interface. One wrote:

*You can improve this website by having us go into more depth of how we handled the situation.*

Referring to the final summary step in StoryPRIME where the interface suggested students to limit their writing to 300 characters, another suggested:

*(Allow) More characters to type your summary with.*

Nevertheless, some of them provided positive feedback about the interface:

*This website is awesome as it is :)*

*Honestly, I don't feel that this website needs any improvement. It is very easy to follow and helped me tell my stories.*

In addition to helping students to write testimonials, StoryPRIME interface seemed to provide them with an

opportunity to reflect and learn from their past experiences. One writer in StoryPRIME condition wrote:

*it was easy and felt nice to describe an incident that helped me become a good student.*

Another wrote:

*...it was an eye opening experience. I learned that I can reflect back on my early high school career in detail.*

Finally, one student wrote:

*I learned that i actually learned from my mistakes as a freshmen to be more successful.*

The StoryPRIME interface helped prevention researchers translate a key component of an existing in-person, evidence-based practice from a public health intervention (ie, eliciting Sources of Strength peer leaders to share their stories) onto a Web-based platform (ie, text messaging), which is a hurdle many evidence-based public health programs face as they seek to increase their reach through technology [2,36,37]. The ability to reliably generate short-form testimonials via StoryPRIME opens up new possibilities for other public health practitioners who wish to draw on peer models to deliver messages through short-form Web-based media, such as text messages and social media microblogs [38]. Ultimately, we plan to use StoryPRIME to elicit strength-based testimonials from members of the target population who are not peer leaders. By translating a school-based technique into a Web- and text messaging-based one, we hope to increase the reach of Text4Strength to an audience broader than the one reached by its parent program, Source of Strength, alone. However, it is worth nothing that in this experiment, we are testing the efficacy of the procedure in generating appealing and relevant testimonials, not the efficacy of the procedure in suicide prevention. The latter will be accomplished in a trial Text4Strength program (currently underway) to which StoryPRIME contributes.

## Limitations

Several limitations should be noted. First, this trial was conducted in 2 small, rural schools in the United States. While we have no reason to expect students from different backgrounds to respond differently from these students, the interface still needs to be tested in a variety of environments and prevention programs. Moreover, the raters in both control and intervention groups might not be representative of all ninth graders in those schools. Second, while StoryPRIME produced more relatable, useful, and likable testimonials, the effect size of the difference between the StoryPRIME and the simple control conditions was not large, so practitioners need to weigh the benefit of using StoryPRIME with the cost of a longer procedure. Finally, a nontrivial portion of the testimonials participants wrote were not interesting or relevant, and could not be used in our program. Thus, we have a procedure that generates high-quality testimonials to choose from, but have not yet arrived at the ideal state in which our interface could elicit youth testimonials ready to be automatically shared without monitoring.

## Comparison With Prior Work

While our work is the first to successfully translate an element of a school-based suicide prevention program onto a Web-based and text messaging-based platform, it builds on previous research on public health related to: (1) text messaging in public health interventions, (2) peer opinion leaders' effectiveness in public health interventions, and (3) other public health efforts to use short form messaging. Previous work with text messaging in nonclinical settings involved smoking cessation [37], child and maternal health, and sexual health [23-25], but ours is the first effort to use texting in population-oriented suicide prevention, as we ultimately aim to do in Text4Strength. Using testimonials created with StoryPRIME in Text4Strength text messages is consistent with "Texting for Public Health Toolkit's" strategy of crafting text messages with "engaging writing devices" [39]. While researchers have shown that peer leaders can effectively influence social norms and behavior in public health contexts, such as suicide prevention interventions [8], and in health care settings [6,7], our work is the first to translate the influence of peer leaders onto Web-based media in order to increase the reach of our intervention. This work is quite different from previous work that translated the content of a curriculum or treatment manual into text [37]. Rather, our work involved translating a complex training and coaching process that occurs in a school-based program into a Web-based interface, ultimately yielding text messages to convey public

health messages. Finally, our work contributes to the growing body of research involving translating evidence-based practices into scalable Web-based formats in order to have a broader public health impact. Programs that involve community members could use a procedure like StoryPRIME to develop compelling testimonials to add to existing public health text messaging programs, such as those used increase adherence to malaria treatment [40,41].

## Conclusions

StoryPRIME serves as an example for translating complex procedures from an in-person evidenced-based program and moving them to a Web-based one in order to scale up technology use. We translated the procedure for collecting peer-to-peer testimonials, a critical component of Sources of Strength. StoryPRIME elicited high school peer leaders to generate engaging testimonials suitable for use in text messaging. The next step is to incorporate StoryPRIME testimonials into Text4Strength, a text messaging extension of Sources of Strength, to determine whether the testimonials are effective in promoting positive coping and emotional skills. Other programs that disseminate peer testimonials in public health messaging can draw on this work to improve the efficiency and effectiveness of collecting brief testimonials for use in short-form, Web-based media, such as text messages and social media microblogs.

## Acknowledgments

We gratefully acknowledge support from the National Institute of Mental Health (K23MH101449). The authors wish to thank Nathaniel Kerr for coordinating aspects of this study, Bill Morrison for his help with manuscript preparation, and Mariya Petrova for her input in translating Sources of Strength training into discrete steps for StoryPRIME.

## Authors' Contributions

Phyo Thiha was the primary designer and programmer of the StoryPRIME interface, contributed to school-based fieldwork, designed and conducted the randomized controlled experiment, and had primary responsibility over manuscript development. Anthony R. Pisani, contributed to the design of StoryPRIME, led all school-based fieldwork, authored major sections of the manuscript, and advised Mr. Thiha on all aspects of the study. Kunali Gurditta contributed to manuscript development and preparation. Erin Cherry contributed to initial design process of StoryPRIME and helped to edit previous versions of this manuscript. Derick R. Peterson designed and conducted all statistical analyses. Henry Kautz provided conceptual input into the design of StoryPRIME and the randomized controlled experiment to test its efficacy. Peter A. Wyman assisted the team in translating school-based procedures from Sources of Strength into the online StoryPRIME interface and provided feedback on the manuscript.

## Conflicts of Interest

None declared.

## References

1. Norman CD, Haresign H, Mehling C, Bloomberg H. Exploring the feasibility and potential of virtual panels for soliciting feedback on nutrition education materials: a proof-of-concept study. *JMIR Public Health Surveill* 2016;2:e18 [FREE Full text] [doi: [10.2196/publichealth.5134](https://doi.org/10.2196/publichealth.5134)] [Medline: [27227153](https://pubmed.ncbi.nlm.nih.gov/27227153/)]
2. Rempel GR, Ballantyne RT, Magill-Evans J, Nicholas DB, Mackie AS. Texting teens in transition: the use of text messages in clinical intervention research. *JMIR Mhealth Uhealth* 2014;2:e45 [FREE Full text] [doi: [10.2196/mhealth.3232](https://doi.org/10.2196/mhealth.3232)] [Medline: [25379624](https://pubmed.ncbi.nlm.nih.gov/25379624/)]
3. Sze YY, Daniel TO, Kilanowski CK, Collins RL, Epstein LH. Web-based and mobile delivery of an episodic future thinking intervention for overweight and obese families: a feasibility study. *JMIR Mhealth Uhealth* 2015;3:e97 [FREE Full text] [doi: [10.2196/mhealth.4603](https://doi.org/10.2196/mhealth.4603)] [Medline: [26678959](https://pubmed.ncbi.nlm.nih.gov/26678959/)]



4. Sikkema KJ, Kelly JA, Winett RA, Solomon LJ, Cargill VA, Roffman RA, et al. Outcomes of a randomized community-level HIV prevention intervention for women living in 18 low-income housing developments. *Am J Public Health* 2000;90:57-63. [Medline: [10630138](#)]
5. Valente TW, Chou CP, Pentz MA. Community coalitions as a system: effects of network change on adoption of evidence-based substance abuse prevention. *Am J Public Health* 2007;97:880-886. [doi: [10.2105/AJPH.2005.063644](#)] [Medline: [17329667](#)]
6. Schneider JA, Cornwell B, Ostrow D, Michaels S, Schumm P, Laumann EO, et al. Network mixing and network influences most linked to HIV infection and risk behavior in the HIV epidemic among black men who have sex with men. *Am J Public Health* 2013;103:e28-e36 [FREE Full text] [doi: [10.2105/AJPH.2012.301003](#)] [Medline: [23153147](#)]
7. Schneider JA, Zhou AN, Laumann EO. A new HIV prevention network approach: sociometric peer change agent selection. *Soc Sci Med* 2015;125:192-202 [FREE Full text] [doi: [10.1016/j.socscimed.2013.12.034](#)] [Medline: [24518188](#)]
8. Wyman PA, Brown CH, LoMurray M, Schmeelk-Cone K, Petrova M, Yu Q, et al. An outcome evaluation of the Sources of Strength suicide prevention program delivered by adolescent peer leaders in high schools. *Am J Public Health* 2010;100:1653-1661. [doi: [10.2105/AJPH.2009.190025](#)] [Medline: [20634440](#)]
9. Bock BC, Rosen RK, Barnett NP, Thind H, Walaska K, Foster R, et al. Translating behavioral interventions onto mHealth platforms: developing text message interventions for smoking and alcohol. *JMIR Mhealth Uhealth* 2015;3:e22 [FREE Full text] [doi: [10.2196/mhealth.3779](#)] [Medline: [25714907](#)]
10. Christakis NA, Fowler JH. The collective dynamics of smoking in a large social network. *N Engl J Med* 2008;358:2249-2258 [FREE Full text] [doi: [10.1056/NEJMs0706154](#)] [Medline: [18499567](#)]
11. Pisani AR, Wyman PA, Petrova M, Schmeelk-Cone K, Goldston DB, Xia Y, et al. Emotion regulation difficulties, youth-adult relationships, and suicide attempts among high school students in underserved communities. *J Youth Adolesc* 2013;42:807-820 [FREE Full text] [doi: [10.1007/s10964-012-9884-2](#)] [Medline: [23666604](#)]
12. Petrova M, Wyman PA, Schmeelk-Cone K, Pisani AR. Positive-themed suicide prevention messages delivered by adolescent peer leaders: proximal impact on classmates' coping attitudes and perceptions of adult support. *Suicide Life Threat Behav* 2015;45:651-663. [doi: [10.1111/sltb.12156](#)] [Medline: [25692382](#)]
13. LoMurray M. Sources of Strength facilitators guide: suicide prevention peer gatekeeper training. Bismarck, ND: The North Dakota Suicide Prevention Project; 2005. URL: <https://sourcesofstrength.org/> [accessed 2016-10-29] [WebCite Cache ID [6ld2vcWxw](#)]
14. Forbes E, Dahl R. Neural systems of positive affect: relevance to understanding child and adolescent depression? *Dev Psychopathol* 2005;17:827-850 [FREE Full text] [doi: [10.1017/S095457940505039X](#)] [Medline: [16262994](#)]
15. Gratz KL, Roemer L. Multidimensional assessment of emotion regulation and dysregulation: development, factor structure, and initial validation of the difficulties in emotion regulation scale. *J Psychopathol Behav Assess* 2004;26:41-54. [doi: [10.1023/B:JOBA.0000007455.08539.94](#)]
16. Jacobson CM, Marrocco F, Kleinman M, Gould MS. Restrictive emotionality, depressive symptoms, and suicidal thoughts and behaviors among high school students. *J Youth Adolesc* 2011;40:656-665. [doi: [10.1007/s10964-010-9573-y](#)] [Medline: [20661633](#)]
17. Neumann A, van Lier PAC, Gratz KL, Koot HM. Multidimensional assessment of emotion regulation difficulties in adolescents using the Difficulties in Emotion Regulation Scale. *Assessment* 2010;17:138-149. [doi: [10.1177/1073191109349579](#)] [Medline: [19915198](#)]
18. Lenhart A. Teens, smartphones & texting. Pew Internet & American Life Project 2012 [FREE Full text]
19. Aguilera A, Muñoz RF. Text messaging as an adjunct to CBT in low-income populations: a usability and feasibility pilot study. *Prof Psychol Res Pr* 2011;42:472-478 [FREE Full text] [doi: [10.1037/a0025499](#)] [Medline: [25525292](#)]
20. Cole-Lewis H, Kershaw T. Text messaging as a tool for behavior change in disease prevention and management. *Epidemiol Rev* 2010;32:56-69 [FREE Full text] [doi: [10.1093/epirev/mxq004](#)] [Medline: [20354039](#)]
21. Arora S, Peters AL, Agy C, Menchine M. A mobile health intervention for inner city patients with poorly controlled diabetes: proof-of-concept of the TExT-MED program. *Diabetes Technol Ther* 2012;14:492-496. [doi: [10.1089/dia.2011.0252](#)] [Medline: [22524591](#)]
22. Mutsuddi A, Connelly K. Text messages for encouraging physical activity Are they effective after the novelty effect wears off? 2012 Presented at: Pervasive Computing Technologies for Healthcare (PervasiveHealth), 2012 6th International Conference; May 21-24, 2012; San Diego, USA. p. 21-24 URL: <http://ieeexplore.ieee.org/document/6240360/> [doi: [10.4108/icst.pervasivehealth.2012.248715](#)]
23. Daniels A, Taylor S, Post S, Pilsner A, Hunt Y, Augustson E. Tech to treat: the smokefree teen approach to cessation. 2012 Presented at: 2012 National Conference on Health Communication Marketing and Media; August 7-9, 2012; Atlanta, GA p. 7-9 URL: <https://cdc.confex.com/cdc/nphic12/webprogram/Paper31682.html>
24. Parker RM, Dmitrieva E, Frolov S, Gazmararian JA. Text4baby in the United States and Russia: an opportunity for understanding how mHealth affects maternal and child health. *J Health Commun* 2012;Suppl 1:30-36. [doi: [10.1080/10810730.2011.649162](#)] [Medline: [22548596](#)]
25. Levine D, McCright J, Dobkin L, Woodruff AJ, Klausner JD. SEXINFO: a sexual health text messaging service for San Francisco youth. *Am J Public Health* 2008;98:393-395. [doi: [10.2105/AJPH.2007.110767](#)] [Medline: [18235068](#)]



26. Results from the 2012 National Survey on Drug Use and Health: Summary of National Findings. US Department of Health and Human Services. 2013 Dec. URL: [http://archive.samhsa.gov/data/NSDUH/2k12MH\\_FindingSandDetTables/2K12MHF/NSDUHmhfr2012.htm](http://archive.samhsa.gov/data/NSDUH/2k12MH_FindingSandDetTables/2K12MHF/NSDUHmhfr2012.htm) [accessed 2016-10-29] [WebCite Cache ID 6ld1zMBU]
27. Casey BJ, Jones RM, Levita L, Libby V, Pattwell SS, Ruberry EJ, et al. The storm and stress of adolescence: insights from human imaging and mouse genetics. *Dev Psychobiol* 2010;52:225-235 [FREE Full text] [doi: [10.1002/dev.20447](https://doi.org/10.1002/dev.20447)] [Medline: [20222060](https://pubmed.ncbi.nlm.nih.gov/20222060/)]
28. Hare TA, Tottenham N, Galvan A, Voss HU, Glover GH, Casey BJ. Biological substrates of emotional reactivity and regulation in adolescence during an emotional go-nogo task. *Biol Psychiatry* 2008;63:927-934 [FREE Full text] [doi: [10.1016/j.biopsych.2008.03.015](https://doi.org/10.1016/j.biopsych.2008.03.015)] [Medline: [18452757](https://pubmed.ncbi.nlm.nih.gov/18452757/)]
29. Somerville LH, Casey BJ. Developmental neurobiology of cognitive control and motivational systems. *Curr Opin Neurobiol* 2010;20:236-241 [FREE Full text] [doi: [10.1016/j.conb.2010.01.006](https://doi.org/10.1016/j.conb.2010.01.006)] [Medline: [20167473](https://pubmed.ncbi.nlm.nih.gov/20167473/)]
30. Kelly A M Clare, Di MA, Uddin LQ, Shehzad Z, Gee DG, Reiss PT, et al. Development of anterior cingulate functional connectivity from late childhood to early adulthood. *Cereb Cortex* 2009;19:640-657 [FREE Full text] [doi: [10.1093/cercor/bhn117](https://doi.org/10.1093/cercor/bhn117)] [Medline: [18653667](https://pubmed.ncbi.nlm.nih.gov/18653667/)]
31. Wilson BJ. Designing media messages about health and nutrition: what strategies are most effective? *J Nutr Educ Behav* 2007;39(Suppl):S13-S19. [doi: [10.1016/j.jneb.2006.09.001](https://doi.org/10.1016/j.jneb.2006.09.001)] [Medline: [17336800](https://pubmed.ncbi.nlm.nih.gov/17336800/)]
32. Lim MSC, Wright C, Hellard ME. The medium and the message: fitting sound health promotion methodology into 160 characters. *JMIR Mhealth Uhealth* 2014;2:e40 [FREE Full text] [doi: [10.2196/mhealth.3888](https://doi.org/10.2196/mhealth.3888)] [Medline: [25367387](https://pubmed.ncbi.nlm.nih.gov/25367387/)]
33. Muller M. Participatory design: the third space in HCI. In: Sears A, Jacko J, editors. *Human-Computer Interaction: Development Process*. Boca Raton, FL: CRC Press; 2009:165-185.
34. Lancaster GA, Dodd S, Williamson PR. Design and analysis of pilot studies: recommendations for good practice. *J Eval Clin Pract* 2004;10:307-312. [doi: [10.1111/j.2002.384.doc.x](https://doi.org/10.1111/j.2002.384.doc.x)] [Medline: [15189396](https://pubmed.ncbi.nlm.nih.gov/15189396/)]
35. Akaliyski P. How can I justify the use of statistical significance at the 10%?. URL: [https://www.researchgate.net/post/How\\_can\\_I\\_justify\\_the\\_use\\_of\\_statistical\\_significance\\_at\\_the\\_10](https://www.researchgate.net/post/How_can_I_justify_the_use_of_statistical_significance_at_the_10) [accessed 2016-09-14] [WebCite Cache ID 6kW4XYCA]
36. Watts S, Mackenzie A, Thomas C, Griskaitis A, Mewton L, Williams A, et al. CBT for depression: a pilot RCT comparing mobile phone vs. computer. *BMC Psychiatry* 2013;13:49 [FREE Full text] [doi: [10.1186/1471-244X-13-49](https://doi.org/10.1186/1471-244X-13-49)] [Medline: [23391304](https://pubmed.ncbi.nlm.nih.gov/23391304/)]
37. Bock B, Heron K, Jennings E, Morrow K, Cobb V, Magee J, et al. A text message delivered smoking cessation intervention: the initial trial of TXT-2-Quit: randomized controlled trial. *JMIR Mhealth Uhealth* 2013;1:e17 [FREE Full text] [doi: [10.2196/mhealth.2522](https://doi.org/10.2196/mhealth.2522)] [Medline: [25098502](https://pubmed.ncbi.nlm.nih.gov/25098502/)]
38. Young SD, Cumberland WG, Lee S, Jaganath D, Szekeres G, Coates T. Social networking technologies as an emerging tool for HIV prevention: a cluster randomized trial. *Ann Intern Med* 2013;159:318-324 [FREE Full text] [doi: [10.7326/0003-4819-159-5-201309030-00005](https://doi.org/10.7326/0003-4819-159-5-201309030-00005)] [Medline: [24026317](https://pubmed.ncbi.nlm.nih.gov/24026317/)]
39. Karasz H, Bosslet L. Northwest Center for Public Health Practice and Seattle and King County Public Health. 2015. Texting for public health toolkit: crafting messages URL: <http://www.nwcphp.org/docs/sms-toolkit/messages/index.htm> [accessed 2016-06-15] [WebCite Cache ID 6iI3g73jr]
40. Jones COH, Wasunna B, Sudoi R, Githinji S, Snow RW, Zurovac D. "Even if you know everything you can forget": health worker perceptions of mobile phone text-messaging to improve malaria case-management in Kenya. *PLoS One* 2012;7:e38636 [FREE Full text] [doi: [10.1371/journal.pone.0038636](https://doi.org/10.1371/journal.pone.0038636)] [Medline: [22719911](https://pubmed.ncbi.nlm.nih.gov/22719911/)]
41. Zurovac D, Sudoi RK, Akhwale WS, Ndiritu M, Hamer DH, Rowe AK, et al. The effect of mobile phone text-message reminders on Kenyan health workers' adherence to malaria treatment guidelines: a cluster randomised trial. *Lancet* 2011;378:795-803 [FREE Full text] [doi: [10.1016/S0140-6736\(11\)60783-6](https://doi.org/10.1016/S0140-6736(11)60783-6)] [Medline: [21820166](https://pubmed.ncbi.nlm.nih.gov/21820166/)]

---

## Abbreviations

**HCI:** human-computer interaction

**HIV:** human immunodeficiency virus

**MTurk:** Mechanical Turk

**StoryPRIME:** stories of personal resilience in managing emotion

---

*Edited by G Eysenbach; submitted 15.06.16; peer-reviewed by C Wright, J Braciszewski, S Berrouiguet, W van Ballegooijen, L Wozney; comments to author 11.07.16; revised version received 20.09.16; accepted 12.10.16; published 09.11.16.*

*Please cite as:*

*Thiha P, Pisani AR, Gurditta K, Cherry E, Peterson DR, Kautz H, Wyman PA*

*Efficacy of Web-Based Collection of Strength-Based Testimonials for Text Message Extension of Youth Suicide Prevention Program: Randomized Controlled Experiment*

*JMIR Public Health Surveill 2016;2(2):e164*

URL: <http://publichealth.jmir.org/2016/2/e164/>

doi: [10.2196/publichealth.6207](https://doi.org/10.2196/publichealth.6207)

PMID: [27829575](https://pubmed.ncbi.nlm.nih.gov/27829575/)

©Phyo Thiha, Anthony R Pisani, Kunali Gurditta, Erin Cherry, Derick R Peterson, Henry Kautz, Peter A Wyman. Originally published in JMIR Public Health and Surveillance (<http://publichealth.jmir.org>), 09.11.2016. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Public Health and Surveillance, is properly cited. The complete bibliographic information, a link to the original publication on <http://publichealth.jmir.org>, as well as this copyright and license information must be included.

Original Paper

# Perceptions Toward a Smoking Cessation App Targeting LGBTQ+ Youth and Young Adults: A Qualitative Framework Analysis of Focus Groups

N Bruce Baskerville<sup>1</sup>, MHA, PhD; Darly Dash<sup>1</sup>, MSc; Katy Wong<sup>1</sup>, MSc; Alanna Shuh<sup>1</sup>, BSc; Aneta Abramowicz<sup>1</sup>, MSc

Propel Centre for Population Health Impact, Applied Health Sciences, University of Waterloo, Waterloo, ON, Canada

**Corresponding Author:**

N Bruce Baskerville, MHA, PhD  
Propel Centre for Population Health Impact  
Applied Health Sciences  
University of Waterloo  
200 University Ave West  
Waterloo, ON, N2L 3G1  
Canada  
Phone: 1-519-888-4567 ext. 35236  
Fax: 1-519-746-8171  
Email: [nbbaskerville@uwaterloo.ca](mailto:nbbaskerville@uwaterloo.ca)

## Abstract

**Background:** The prevalence of smoking among lesbian, gay, bisexual, trans, queer, and other sexual minority (LGBTQ+) youth and young adults (YYA) is significantly higher compared with that among non-LGBTQ+ persons. However, in the past, interventions were primarily group cessation classes that targeted LGBTQ+ persons of all ages. mHealth interventions offer an alternate and modern intervention platform for this subpopulation and may be of particular interest for young LGBTQ+ persons.

**Objective:** This study explored LGBTQ+ YYA (the potential users') perceptions of a culturally tailored mobile app for smoking cessation. Specifically, we sought to understand what LGBTQ+ YYA like and dislike about this potential cessation tool, along with how such interventions could be improved.

**Methods:** We conducted 24 focus groups with 204 LGBTQ+ YYA (aged 16-29 years) in Toronto and Ottawa, Canada. Participants reflected on how an app might support LGBTQ+ persons with smoking cessation. Participants indicated their feelings, likes and dislikes, concerns, and additional ideas for culturally tailored smoking cessation apps. Framework analysis was used to code transcripts and identify the overarching themes.

**Results:** Study findings suggested that LGBTQ+ YYA were eager about using culturally tailored mobile apps for smoking cessation. Accessibility, monitoring and tracking, connecting with community members, tailoring, connecting with social networks, and personalization were key reasons that were valued for a mobile app cessation program. However, concerns were raised about individual privacy and that not all individuals had access to a mobile phone, users might lose interest quickly, an app would need to be marketed effectively, and app users might cheat and lie about progress to themselves. Participants highlighted that the addition of distractions, rewards, notifications, and Web-based and print versions of the app would be extremely useful to mitigate some of their concerns.

**Conclusions:** This study provided insight into the perspectives of LGBTQ+ YYA on a smoking cessation intervention delivered through a mobile app. The findings suggested a number of components of a mobile app that were valued and those that were concerning, as well as suggestions on how to make a mobile app cessation program successful. App development for this subpopulation should take into consideration the opinions of the intended users and involve them in the development and evaluation of mobile-based smoking cessation programs.

(*JMIR Public Health Surveill* 2016;2(2):e165) doi:[10.2196/publichealth.6188](https://doi.org/10.2196/publichealth.6188)

**KEYWORDS**

qualitative research; focus groups; mobile applications; telemedicine; sexuality; smoking cessation; smartphone

## Introduction

Significant reductions in tobacco prevalence and use have been achieved; however, tobacco use continues to be the leading cause of preventable disease, disability, and death in Canada. In 2014, 18.1% of Canadians aged 12 years and older were daily or occasional smokers [1] but this prevalence is not equal across all subpopulations. Smoking prevalence is particularly high in lesbian, gay, bisexual, trans, queer, and other sexual minority (LGBTQ+) persons with estimates ranging between 24% and 45% across different sexual orientation and gender identity groups [2]. There are even higher rates among LGBTQ+ youth and young adults (YYA) [2]. According to the 2013-2014 Canadian Community Health Survey (CCHS), 34.0% of homosexuals (interpret with caution due to marginal sampling variability) and 35.1% of bisexuals among 18- to 24-year-olds report smoking daily or occasionally compared with 23.3% of heterosexuals [3]. Further, 22% of high school-aged adolescents who identify as lesbian, gay, or bisexual report daily cigarette use compared with 11% of non-LGBTQ+ adolescents [4].

There are several suggested factors that contribute to high smoking rates among LGBTQ+ persons. These include minority stress, victimization, discrimination, harassment, abuse, mental health, targeted marketing by the tobacco industry, frequenting bars and nightclubs, other substance use, higher rates of stress, depression, and alcohol use, and low socioeconomic status [5-9].

Remafedi and Carol (2005) interviewed LGBTQ+ youth who highlighted the need for programs to be culturally specific and targeted toward youth, and that programming should involve LGBTQ+ in planning and implementation [10]. However, to date, there are few smoking cessation interventions and studies that specifically target LGBTQ+ YYA [11,12]. The majority of interventions are geared toward LGBTQ+ individuals of all ages and involve group cessation counseling or broader social marketing campaigns, where the interventions have been either tailored to LGBTQ+ people or assessed for effectiveness within the community. A recent review highlighted that group counseling programs showed evidence for effectiveness but had low reach in the target population [11]. Further, media and social marketing campaigns were found to be successfully implemented and increased media coverage and outreach to LGBTQ+ organizations; however, as they are primarily descriptive studies, there is limited evidence of impact [11].

Recently, digital technologies have shown promise [13,14] and offer potential avenues to deliver a smoking cessation program to LGBTQ+ YYA. For example, the mobile phone market is growing rapidly with statistics from the United States showing that mobile phone ownership in adults was 68% in 2015 [15]. In young persons (aged 18-29 years), ownership is as high as 86% and is closely reaching saturation [15]. Similar to the United States, Canadian data from 2015 indicate that, overall, 68% of adults own a mobile phone [16], while 2013 data indicate that 86% of 18- to 34-year-olds own a mobile phone [17]. Of interesting note, minorities and those with lower income levels are more likely to access the Internet primarily through mobile phones [18]. Given continual growing rates of mobile phone ownership across the population, mobile health (mHealth)

interventions offer a potential method of engaging YYA who identify as LGBTQ+.

There are numerous benefits to delivering a smoking cessation program through a mobile phone app compared with traditional approaches. These include low cost of the intervention, accessibility, portability, personalization, self-monitoring capabilities, location-determining sensors, access to the Internet, and ability to connect with social networking platforms [19-21].

There is a growing body of evidence supporting mobile phone-based interventions for smoking cessation, although studies have primarily evaluated text messaging interventions for smoking cessation [13,20]. Research on mobile apps has been limited to small pilot studies [22-24] and trial protocols currently in progress [25-29]. Smoking cessation apps are also being developed and delivered rapidly as hundreds of apps are available on iPhone and Android platforms. Studies have found that publicly available cessation apps do not conform to established best practice guidelines including the 5A protocol: ask, advise, assess, assist, and arrange follow-up [30-32], do not recommend quitline use or have low rates for recommending approved medications [31-33], and do not take full advantage of interactive and proactive tailoring for the consumer [33]. However, there are no studies reporting whether LGBTQ+ YYA are interested in a mobile phone app for smoking cessation.

We addressed this gap in the literature by focusing on the LGBTQ+ YYA population. This study explored LGBTQ+ YYA (the potential users') perceptions on a culturally tailored mobile app for smoking cessation. Potential users shared their opinions on this hypothetical cessation tool, along with opinions on what content and design features should be included.

## Methods

### Design and Recruitment

We conducted a total of 24 focus groups (FGs) among LGBTQ+ YYA in Toronto (n=18 groups) and Ottawa (n=6 groups) from March to May 2015. Recruitment was done using a variety of methods through purposive and snowball sampling: posting flyers and verbal announcements in spaces frequented by LGBTQ+ people; Facebook posts on LGBTQ+-friendly group pages; paid Facebook advertisements; emails from LGBTQ+ agencies; organizations making calls through social media channels; targeted physical recruitment at bars and nightclubs in Ottawa; and asking participants to refer eligible peers.

At the time of recruitment, interested participants contacted the project coordinator via email and completed an electronic demographic intake questionnaire. Physical questionnaires were available for those unable to complete the electronic form. Eligible participants were 16-29 years old, a part of the LGBTQ+ community, and a current smoker or recent quitter (defined as having not quit for more than 6 months before completing the intake questionnaire). Thus, those who had quit for more than 6 months or those who had never smoked were ineligible to participate. Those who were eligible were triaged into age, city, and LGBTQ+ groups for the purpose of organizing and hosting FGs based on location (ie, Toronto vs Ottawa), age (youth aged 16-18 years vs young adults aged

19-29 years), and gender or sexual identity of participants. The objective was to create FG conditions that were the most conducive for open and safe dialogue among participants (eg, gay youths speaking with other gay youths). The project coordinator confirmed attendance and information about FGs via email and attended all sessions. The attendance rate at the FGs was 74% (204 individuals attended out of the 275 who agreed to attend).

### Focus Group Procedures

Facilitators and note takers (one each per location) were trained in study objectives and the semistructured FG script and protocol. The FG script with questions and description of the mHealth idea can be found in [Multimedia Appendix 1](#).

**Textbox 1.** Mobile app intervention idea for LGBTQ+ youth and young adults.

Do you own a smartphone? Ever play Candy Crush or use Instagram? What if there was an app that could help you quit smoking designed specifically for LGBTQ+ youth and young adults? For example, this quit smoking app would allow you to create an individualized quit plan where you can set a quit date, it would provide feedback on how you're doing, record what triggers you to smoke, and give you tips on how to remain smoke-free, as well as links to counselling services. One of the advantages of the app would be access to a peer support network which would connect you to other LGBTQ+ peers who are also trying to quit or who have already stopped smoking.

The app would be part of a bigger social media campaign that would include a webpage, Facebook page, YouTube videos, and Twitter feed with access to more detailed educational resources about smoking and quitting [eg, nicotine replacement therapy, like gum or the patch]. LGBTQ+ role models would promote the campaign.

The FGs contained 3-17 participants and lasted for 90 minutes. Participants were remunerated with a Can \$50 cash incentive. All FGs were audio-recorded and professionally transcribed. This study was approved by a University of Waterloo Research Ethics Committee.

### Data Analysis

We qualitatively analyzed the data using a framework analysis technique that incorporated the stages of familiarization, identification of a thematic framework, indexing, charting, mapping, and interpretation [34]. Coding categories were developed and entered into the NVivo 11 (QSR International) software. To validate coding, 2 authors (AS and KW) independently coded 6 FG transcripts and then compared for consistency. Any discrepancies in coding were discussed and resolved with the principal investigator (NB). In this way, each author was able to critically challenge one another on differing perspectives and any potential biases.

After coding of the initial transcripts, a thematic framework was developed from the a priori issues and emerging codes by generating major themes and subthemes in relation to the original FG questions. Associated responses were categorized iteratively. To maintain the context of FG participant responses, they were listed under the questions from which they were derived and then categorized separately as a type of response. Throughout the coding or indexing process, regular meetings were held between 3 of the authors to discuss and refine the emerging thematic framework. Indexing was accomplished by coding each FG in NVivo 11, with reliability checked by a third author (DD) through review of the NVivo 11 file. At the final stage, the original responses were charted and grouped according to the finalized themes and subthemes. Member checking was completed with 14 participants to confirm findings. The final stage of mapping and interpretation involved comparing and

Facilitators and note takers identified as part of the community. Participants provided written consent. FGs were conducted in LGBTQ+ community safe spaces. Participants were provided with the mobile app intervention idea, as part of a broader program of research to explore potential interventions for LGBTQ+ YYA. The intervention idea was handed out on paper to each participant who were also provided with a pen and notepad to write down their thoughts about the LGBTQ+-specific mobile smoking cessation app intervention ([Textbox 1](#)). They were then asked to verbally describe and share their thoughts and feelings with the group. This included sharing perceived usefulness, relevance, likes and dislikes, concerns, and ideas or suggestions for further improvement of this intervention idea.

contrasting the responses to search for patterns. Responses were grouped according to the final themes and subthemes that emerged from the data. Saturation of findings occurred after analysis of the fifteenth FG transcript. Representative quotes were selected from the FG responses to illustrate key themes and subthemes.

We used the Consolidated Criteria for Reporting Qualitative Research guideline statement to assist in the reporting of the study [35].

## Results

### Participants

A total of 204 participants were recruited in 24 FGs with 18 youth (aged 16-17 years) and 186 young adults (aged 19-29 years). We conducted 11 FGs with mixed gender and sexual orientation participants and 13 gender- or sexual orientation-specific groups (3 gay, 3 bisexual, 3 transgender, 2 lesbian, and 2 queer groups).

The majority of participants identified as white (48%, 115/241), female gender (39%, 85/218), bisexual orientation (27%, 57/209), and renting or owning their homes (50%, 118/234), and the majority completed high school (77%, 158/204). Also, 84% (171/204) of individuals were daily or occasional smokers, and 37% (75/204) of individuals reported smoking within 30 minutes after waking. Additional characteristics are depicted in [Multimedia Appendix 2](#).

### Overview

We found that responses toward a culturally tailored smoking cessation app focused on 3 main areas: (1) why a mobile app for cessation would work; (2) concerns over the use of a mobile



app; and (3) suggestions that should be incorporated into a mobile app for this population.

### **Reasons Why a Mobile App Would Work**

The majority of participants in this study overwhelmingly expressed positive feelings toward a mobile smoking cessation app. Many supported this type of intervention and were excited about its potential.

*I'm definitely most likely to use the app. That actually excites me thinking about it because I never really thought about that and people I know would use it – hands down, for sure. [FG02, Gay group participant]*

*I think it would be an awesome idea for me if there were some apps for quitting smoking because I can't leave my phone for one hour. [FG18, Mixed group participant]*

A number of specific factors made a smoking cessation app appealing for LGBTQ+ YYA. First, the most commonly expressed benefit was its potential for accessibility and availability at all times, without requiring a visit to see a counselor, physician, or other health professionals.

*Yes, the (app) idea (is) going to work because I can access it at home as opposed to when I have to go out. I get access at any time, anywhere. I get it when I wake up first thing in the morning (and) before I go to bed. [FG06, Mixed group participant]*

Next, participants liked that apps could provide the mechanism to help them monitor and track their attempts and progress in changing behavior. Individuals reflected that this would allow them to understand their progress and notice patterns. In several instances, participants indicated that the app should be able to track the number of cigarettes consumed, the amount of money someone was saving while not smoking, and how long they had been smoke-free in hours and days.

*I like this idea, I think it's really cool. Having a way to track my smoking because often times I just don't know how much (I smoke); I'm not paying attention to it. But if I had an app that I can check in on or sign into, I can probably keep track of it and eventually help me quit. [FG09, Lesbian group participant]*

*An app that would help me is (one) that I could input on a daily basis how many cigarettes I've taken (per) day and... monitor my own consumption of cigarettes and then wean myself down. Sometimes you don't even realise how much you've smoked. If I had my phone telling me how many I've had, then I'm going to be like, 'oh, okay,' and that would help me keep track. [FG24, Mixed group participant]*

In addition to liking an app due to its tracking and accessibility, LGBTQ+ YYA would like to see a way to connect with others within their community who are also trying to quit smoking. This type of networking system would allow app users to encourage each other, share tips and tricks, and know there are others like them undergoing similar experiences. This could also be an avenue to connect with professionals and counselors,

as this would be an easy way to access these professional resources.

*If it were to heavily play into the community-base where it's not just you on your phone and you get reminders, but if you were able to connect with other people, within a marginalised community, I think that would be great. I think (if) you could specifically reach out to certain groups, you know, trans people supporting trans, people of colour supporting people of colour... I think that would be a great option. [FG21, Mixed group participant]*

*I do like getting access to counselling services because it's so hard to get those resources, maybe there could even be a counsellor linked up in the app. [FG23, Mixed group participant]*

The next positive factor was tailoring of the app to the LGBTQ+ community, and this was considered very important to FG participants. In general, participants felt that an LGBTQ+-specific app would be superior to a general app, as the community would be friendly to those with LGBTQ+ orientation. There could be considerations embedded toward the struggles LGBTQ+ persons may have been through such as victimization. LGBTQ+ role models could also be prominently featured to influence smoke-free behavior.

*Assuming that it's customised to LGBTQ (and) it incorporates the kinds of struggles that we've lived through, it wouldn't be any average quit-smoking app. The fact that it's specific to a community... the fact that it's LGBTQ-specific, that would help us more than if it was just a general quit-smoking app. [FG05, Mixed group participant]*

The fifth factor revolved around apps having the ability to link and share information via other social networks (eg, Facebook, Twitter). Participants felt that integration with such networks was a vital factor, as they would be able to share their experiences to garner additional support. It is important to note that participants felt that linking to a social network should be a choice, not a mandatory feature within the app.

*Integrating it with the social media is definitely a great thing to do because they can always fall back to Facebook, Twitter, etc. And through this, people can get to share their experiences and keep an update and tell whatever experiences they may have to share. So it's like ongoing support. [FG09, Lesbian group participant]*

Another essential feature that participants felt a quit smoking app should possess is the ability for the app to be personalized and customized by the user. For instance, participants shared that they could personalize the quit smoking goal in steps that were achievable to them. Further, the idea of uploading and posting personal pictures was revealed as a motivational tool that could be achieved through customization.

*I think (the) key is that folks can personalize their goal and how they want to record and achieve it and see it monitored. Yes, that's it. [FG21, Mixed group participant]*

Other reasons that were shared for liking a mobile app included that it would make individuals accountable to themselves, there would be no pressure to use it as it is more casual, it is preferable for those who wish for more privacy (as opposed to group counseling sessions), the intervention would appeal to the youth or young adult population, the idea itself is innovative and modern, and many individuals use other apps that are already successful.

### **Concerns Over the Use of a Mobile App for Smoking Cessation**

Even though attitudes toward a culturally tailored quit smoking app were primarily positive, a number of concerns were raised by participants. The most important concern shared was that not all individuals possess a mobile phone, use social networking sites, or use apps.

*I know everybody kind of has smart phones but some people don't and I guess you kind of have to acknowledge that area of our community that doesn't have access to these things. So it completely takes them out of the picture like if you're completely [offline], you're not even aware of it type of thing. [FG09, Lesbian group participant]*

Further, even though this idea was considered to be “awesome,” “exciting,” and “cool,” several participants felt that they might lose interest quickly. This was concerning, as the effectiveness of a quit smoking app on its own would be reduced if users do not access the resource.

*It sounds like a good idea but I think it's, kind of, like, fitness apps or something. They're there, you know; you record the calories you eat for, like, the first few days you download it. And then after, it kind of loses the appeal. It's like any app that you would download on your phone. So I think... it sounds like a good idea. But I think, for most people, it might just lose that interest after a while in having to go in there and, like, record stuff. [FG02, Gay group participant]*

The app market is currently filled with numerous quit smoking apps, some of which are at a significant cost for the user. Participants were concerned that a culturally tailored app would be in competition with other apps, unless properly marketed and shared through appropriate channels for the LGBTQ+ community.

*The only thing with apps though, probably just being cynical again, but I feel like apps, it's like such a saturated environment, there's so many apps out there, I have so many apps on my phone I don't use, like just pages on pages, like I will just scroll through, like I never use them. So I think it would be important that it had that kind of integrity behind it, and a good design behind it, to start out with, because like you were saying, there's so many apps and there's so many repeats of apps. So you would have to get a really good one out on the market first. [FG14, Lesbian group participant]*

Lastly, although the privacy offered by an app was a critical benefit, participants highlighted how this might make it easy to cheat and lie about progress to oneself.

*You could be smoking and posting, oh, I'm not smoking – and you have the cigarette in your mouth still... And when it comes to lying to the app, like, you're only really lying to yourself. That's who you're really lying to, you know? And also, I don't know, I think it would be effective. I would use it. [FG02, Gay group participant]*

*They don't have to be sincere or genuine with the amount of cigarettes they smoked. The time they smoked. How often they smoked or even how often they didn't smoke seeing that you can set up your goal... and then after that goal return back to your old habits. You've accomplished your three-month goal so three months is up now so back to the (old habits). [FG20, Trans group participant]*

### **Suggestions to Incorporate Into a Mobile App**

Participants voiced a number of suggestions that would improve a quit smoking app intervention. Many of these suggestions stemmed from the concerns identified by participants and were offered as solutions to these concerns.

Several participants shared that some method of distraction should be included in the quit smoking app to help deal with cravings. This could be in the form of games or various mind puzzles that keep the smokers' hands occupied.

*If there was a bunch of games on the app that were there to distract you from smoking, (you could) go play five minutes of a quick game instead of smoking. [FG07, Mixed group participant]*

*I think if there were some games on it, because that's one of the things—if my hands aren't busy then I'm going to smoke—so have something kind of fun like that. [FG16, Queer group participant]*

Another key addition proposed was that there should be a feature that provides rewards to the app user. Multiple reward ideas were proposed, such as virtual awards through the app, quit smoking contests using the app with associated prizes, and coupons or other incentives. Participants even commented that organizations who want smokers to quit may be willing to sponsor rewards through the quit smoking app.

*If the quitting process takes a full year, unless you have some contests going on and perks associated with it, it (won't) work. You can get a lot of people to sponsor prizes, like hospitals, doctors' offices, and lotteries that are already trying to get people to quit smoking. There are a lot of organizations that are trying to get people to quit and I think they would support it with some prizes and stuff like that. [FG01, Gay group participant]*

Participants also thought it vital that any quit smoking app tailored to the LGBTQ+ community have a Web-based version or a version that operates even when there is no access to the Internet. A key concern that contributed to this idea was that

many LGBTQ+ individuals are in poverty or homeless, and that there should not be an assumption that all potential users may have access to a mobile phone. It was suggested that the quit smoking app program also have print resources available (eg, pamphlets).

*I feel like there would need to be a website equivalent with it (for) people who don't have access to smartphones but do have access to public libraries. A lot of smokers are LGBTQ and a lot of LGBTQ are in poverty and homeless. The people that you want to access might not be able to access the program.* [FG14, Lesbian group participant]

Other improvements recommended that apps should incorporate a notification feature that reminds individuals to utilize the app and provides encouragement to the user.

*It would be more effective if they send you those notifications every couple (of) hours. So then you have constant updates, and it's a constant reminder. Every time my phone rings or dings I instantly pick it up. Once that would go off, I'd be like, 'oh, okay, my lungs are starting to clear – that's great.* [FG02, Gay group participant]

Further, personalization should allow users to create personal challenges that are relevant and achievable to them (eg, day 1 goal—go for a 20-minute walk around the block). Finally, there is need for the app to be positive, engaging, and cool, and it should be provided at no cost to the user.

Other considerations for the development and implementation of a tailored quit smoking app included a range of design factors that relate to the ease of use and quality of experience of using such an app. Participants indicated that the finished product should: be of high quality and bug free; protect individual app users' privacy at all times; be available for all types of devices, such as Black Berry and not just Apple and Android devices; and lastly, be free of charge to use. Participants noted that ultimately, the effectiveness of such a tool was dependent on the app users' level of dedication and desire to quit: "I don't think any of these ... ideas, would help me quit, or even... (motivate me) to quit smoking." [FG13, Bisexual group participant]

## Discussion

### Principal Findings

This study was the first to explore the perceptions of a large group of urban, Ontario LGBTQ+ YYA smokers and recent quitters about a culturally tailored smoking cessation app. The purpose was to discover what components of a mobile app smoking cessation program may resonate with LGBTQ+ YYA and what concerns and suggestions they may have about cessation apps. This addressed an important gap in the literature, as we sought input from a priority population with high smoking rates.

The overall findings revealed that LGBTQ+ YYA are quite enthusiastic about the role of mobile apps for smoking cessation. We learned that mobile apps could be harnessed for use within

the LGBTQ+ community in a culturally tailored app that was specific and relatable, rather than being geared toward the general population. Other research has found that young persons are cautiously optimistic about the role of apps in health behavior change [19]; however, since that time, the explosion of apps in the market has exponentially increased and gained popularity. More recent research indicates that many current smokers would be interested in using a smoking cessation app [21].

FG participants offered reasons as to why a mobile smoking cessation app would work, including accessibility, portability, tracking and awarding progress, connecting with other community members and social networks, tailoring, notifications, and opportunity for customization. The reasons or preferences for the mobile phone app expressed by participants align with research on persuasive design where it has been shown that self-care technology such as a mobile phone app for smoking cessation should show understanding, persuade users to do the right thing, motivate behavior change, and provide rewards and appraisal for appropriate behavior as well as social support [36,37]. The feedback provided by a mobile phone app has the potential for informing, enabling, motivating, self-regulating, and guiding smokers in their efforts to successfully quit smoking [38].

The preferences for a mobile phone app were often cited in relation to experiences with other smoking cessation interventions that often require in-person meetings or appointments at significant time, travel, and cost to the individual. This finding adds to previous research that indicates that smokers are interested in finding help for quitting in new, more accessible mediums instead of support from quitlines, group cessation classes, or health professionals [39]. Of interest, LGBTQ+ YYA desired communication with other LGBTQ+ persons and community members, over that of health care providers, but through the app medium. However, some participants cited the importance of health professionals and counselors who should be linked with the app, as they are a key resource that can be difficult to access.

We also found that participants placed importance on privacy of their information, as individuals were concerned about protecting their LGBTQ+ identity, smoking status, and quit smoking progress. Although desire for "peer" social support was a highly expressed need from participants, "peer," in this instance, referred to other anonymous LGBTQ+ YYA who are going through the same or similar quit smoking process, rather than their friends and family. Thus, private, as opposed to public, forums or chat spaces were what was desired among this specific group to help them feel safe to openly discuss their quit smoking progress.

Although some research indicates that smokers may not find games to be the most appealing feature [21], in this study, participants felt that gamification was an important form of distraction to the app user as a coping mechanism for cravings. This was critical in the minds of the participants and is supported by evidence demonstrating that assistance with cravings or provision of coping mechanisms such as distractions is central to successfully stopping cigarette addiction [40].



Lee et al (2014) have suggested that nontailored interventions may work as well as tailored for LGBTQ+ smokers [11]. However, FG participants indicated that a mobile phone app tailored specifically for LGBTQ+ was important. Specifically, respondents indicated that it was important that LGBTQ+ were involved in the creation of the app to ensure that it incorporated real-world experiences of the community. Moreover, there was an expressed desire to incorporate LGBTQ+ role models and the potential ability to connect with other LGBTQ+ YYA supports, such as LGBTQ+ counselors and other LGBTQ+ YYA peers (eg, trans youth supporting trans youth) who are going through the quit smoking process. Participants indicated that these specific LGBTQ+ tailored components would have a lot of influence and provide more incentive for use than a general quit smoking app. Considerable public health research has demonstrated the effectiveness of both targeted health promotion messages for population subgroups and individually tailored programs [41,42]. For example, personalized feedback can be adjusted to participants' efficacy level and the unique impediments in their lives, and as a consequence, a smoking cessation app that can be tailored to LGBTQ+ smokers will encourage adoption and reach of the intervention as well as improved cessation outcomes. In addition, recent evidence based on the content analysis of mobile phone apps for smoking cessation suggests that tailoring of apps is positively related to user-rated quality [33].

Participants recognized the importance of social support in quitting smoking and commented on how the app could connect with others within their community. The personal connections that can be formed through a mobile phone app or social media can be viewed as social support, and social support is effective in helping people quit smoking [43]. Further, a community of LGBTQ+ app users may provide the opportunity for smokers to work together to quit smoking and improve the quality of their lives, as research on the effect of social media has revealed that sites, such as Facebook, can be harnessed for supporting young adults who are trying to quit smoking or have become smoke-free [44].

Finally, evidence in support of smoking cessation mobile phone apps is nascent. However, evidence is quickly building with a number of recent study protocols to test the effectiveness of mobile phone apps for smoking cessation as well as reviews of app content for evidence-based behavior change strategies [22-29,31-33].

### Limitations

This research had several limitations. The perceptions shared by FG participants were discussions of a hypothetical app use or experiences with previous app use situated in a particular time frame. It is unknown whether these speculations may change when an actual app is tested and used by the target population. Additionally, we did not probe participants on specific app features but allowed discussion to evolve organically so that participants could identify what was most important to them. However, participants in various FGs may

have initiated conversation on features that others never spoke of. Third, we were unable to obtain the same number of participants within lesbian, gay, bisexual, transgender, and queer groups, and had to create mixed FG sessions. In addition, analysis for this paper was not conducted separately for each group, so it is unknown whether perceptions vary between LGBTQ+ groups. Further, more young adults than youth partook in this study. Finally, it is unknown whether the LGBTQ+ YYA in our sample will generalize to LGBTQ+ smokers in other communities and countries, as the context of our sample came from urban areas where services are typically available for those identifying as LGBTQ+. However, due to the paucity of research on YYA who identify as LGBTQ+, this research sheds light on the perceptions and opinions of this subpopulation group.

### Future Research

Although the consultation with LGBTQ+ young adult smokers shows promise [21-24], future research is needed to determine whether mHealth interventions can help individuals induce behavior change [45] such as quitting smoking. The LGBTQ+ YYA in this study were very interested in an app, so further research is needed to determine whether this population can benefit from a mobile-based smoking cessation intervention with social media connections. It would be useful to study actual app use and whether the findings that emerged from this study occur in a real-life situation. LGBTQ+ YYA should also be involved in the development of a mobile app through participatory research. Finally, additional research is needed to investigate the perceptions of various sexual and gender minorities to determine whether the differences between groups require tailoring specific to each subgroup. Furthermore, in addition to the need for assessing the effectiveness of apps in terms of abstinence from smoking, other important outcome measures such as increased quit attempts should be assessed, as Zhu et al (2012) have demonstrated that increased numbers of quit attempts are key to eventually quitting [46]. Once demonstrated as cost-effective and for whom, policy makers can determine how best to integrate mHealth and social media interventions within comprehensive tobacco cessation systems [31,33].

### Conclusions

mHealth interventions are increasingly popular for research and are a current, contemporary tool with potential to reach many individuals worldwide. Given the lack of consideration in the literature of LGBTQ+ YYA cessation interventions, this formative research contributes to knowledge in regards to what this population sees as important and beneficial in quitting smoking. We found that a mobile app intervention was viewed positively as a potential smoking cessation tool, and this highlights the need to develop such an app with LGBTQ+ YYA involved in the process. In addition to summative evaluation of a mobile phone cessation app's effectiveness for LGBTQ+ YYA, additional formative research on the development and implementation of a mobile phone app is necessary to address knowledge gaps.

## Acknowledgments

This work was supported by the Ontario Ministry of Health, Health Services Research Fund (grant #06696), and the Canadian Cancer Society Research Initiative (grant #701019).

Analysis of the Canadian Community Health Survey was supported by funds to the Canadian Research Data Centre Network from the Social Sciences and Humanities Research Council, the Canadian Institute for Health Research, the Canadian Foundation for Innovation, and Statistics Canada. Although estimates are based on data from Statistics Canada, the opinions expressed do not represent the views of Statistics Canada.

We thank Amer Esmail and Anna Travers from Sherbourne Health Centre, Toronto, Ontario, and Rainbow Health Ontario for assistance and support in conducting the research. We would also like to thank the Editor of this journal and the referees for their helpful comments and suggestions.

## Authors' Contributions

NBB led the conceptualization and design of the study. KW, AS, and AA contributed to the design and implementation of the study, while KW and AS analyzed the results. NBB, KW, and DD contributed to the interpretation of the results. DD and NBB drafted the manuscript. NBB, KW, AS, AA, and DD critically revised the manuscript. NBB supervised the study.

## Conflicts of Interest

None declared.

## Multimedia Appendix 1

LGBTQ+ project: focus group script and questions.

[[PDF File \(Adobe PDF File\), 43KB - publichealth\\_v2i2e165\\_app1.pdf](#)]

## Multimedia Appendix 2

Demographic and smoking characteristics of focus group participants.

[[PDF File \(Adobe PDF File\), 45KB - publichealth\\_v2i2e165\\_app2.pdf](#)]

## References

1. Statistics Canada. Smokers, by sex, provinces and territories. Ottawa, ON: Statistics Canada; 2016. URL: <http://www.statcan.gc.ca/tables-tableaux/sum-som/101/cst01/health74b-eng.htm> [accessed 2016-03-11] [WebCite Cache ID 6fwCStNJF]
2. Clarke MP, Coughlin JR. Prevalence of smoking among the lesbian, gay, bisexual, transsexual, transgender and queer (LGBTQ) subpopulations in Toronto — the Toronto Rainbow Tobacco Survey (TRTS). *Can J Public Health* 2012;103(2):132-136. [Medline: [22530537](#)]
3. Health Statistics Division, Statistics Canada. Canadian Community Health Survey (CCHS) Annual Component 2013-2014 Microdata file. Ottawa, ON: Statistics Canada; 2015.
4. Azagba S, Asbridge M, Langille D, Baskerville B. Disparities in tobacco use by sexual orientation among high school students. *Prev Med* 2014 Dec;69:307-311. [doi: [10.1016/j.ypmed.2014.07.042](#)] [Medline: [25450168](#)]
5. Blossich J, Lee JG, Horn K. A systematic review of the aetiology of tobacco disparities for sexual minorities. *Tob Control* 2013 Mar;22(2):66-73 [FREE Full text] [doi: [10.1136/tobaccocontrol-2011-050181](#)] [Medline: [22170335](#)]
6. Gamarel KE, Mereish EH, Manning D, Iwamoto M, Operario D, Nemoto T. Minority stress, smoking patterns, and cessation attempts: findings from a community-sample of transgender women in the San Francisco Bay area. *Nicotine Tob Res* 2016 Mar;18(3):306-313. [doi: [10.1093/ntr/ntv066](#)] [Medline: [25782458](#)]
7. Newcomb ME, Heinz AJ, Birkett M, Mustanski B. A longitudinal examination of risk and protective factors for cigarette smoking among lesbian, gay, bisexual, and transgender youth. *J Adolesc Health* 2014 May;54(5):558-564 [FREE Full text] [doi: [10.1016/j.jadohealth.2013.10.208](#)] [Medline: [24388111](#)]
8. Remafedi G. Lesbian, gay, bisexual, and transgender youths: who smokes, and why? *Nicotine Tob Res* 2007 Jan;9(Suppl 1):S65-S71. [doi: [10.1080/14622200601083491](#)] [Medline: [17365728](#)]
9. Youatt EJ, Johns MM, Pingel ES, Soler JH, Bauermeister JA. Exploring young adult sexual minority women's perspectives on LGBTQ smoking. *J LGBT Youth* 2015;12(3):323-342 [FREE Full text] [doi: [10.1080/19361653.2015.1022242](#)] [Medline: [26508993](#)]
10. Remafedi G, Carol H. Preventing tobacco use among lesbian, gay, bisexual, and transgender youths. *Nicotine Tob Res* 2005 Apr;7(2):249-256. [doi: [10.1080/14622200500055517](#)] [Medline: [16036282](#)]



11. Lee JG, Matthews AK, McCullen CA, Melvin CL. Promotion of tobacco use cessation for lesbian, gay, bisexual, and transgender people: a systematic review. *Am J Prev Med* 2014 Dec;47(6):823-831 [FREE Full text] [doi: [10.1016/j.amepre.2014.07.051](https://doi.org/10.1016/j.amepre.2014.07.051)] [Medline: [25455123](https://pubmed.ncbi.nlm.nih.gov/25455123/)]
12. Burkhalter JE. Smoking in the LGBT community. In: Boehmer U, Elk R, editors. *Cancer and the LGBT Community*. Springer International Publishing; 2015: 63-80 URL: [http://link.springer.com/chapter/10.1007/978-3-319-15057-4\\_5](http://link.springer.com/chapter/10.1007/978-3-319-15057-4_5) [accessed 2016-11-08] [WebCite Cache ID [6lruCw0y6](https://www.webcitation.org/6lruCw0y6)]
13. Whittaker R, McRobbie H, Bullen C, Borland R, Rodgers A, Gu Y. Mobile phone-based interventions for smoking cessation. *Cochrane Database Syst Rev* 2012;11:CD006611. [doi: [10.1002/14651858.CD006611.pub3](https://doi.org/10.1002/14651858.CD006611.pub3)] [Medline: [23152238](https://pubmed.ncbi.nlm.nih.gov/23152238/)]
14. Free C, Phillips G, Galli L, Watson L, Felix L, Edwards P, et al. The effectiveness of mobile-health technology-based health behaviour change or disease management interventions for health care consumers: a systematic review. *PLoS Med* 2013;10(1):e1001362 [FREE Full text] [doi: [10.1371/journal.pmed.1001362](https://doi.org/10.1371/journal.pmed.1001362)] [Medline: [23349621](https://pubmed.ncbi.nlm.nih.gov/23349621/)]
15. Anderson M. *Technology Device Ownership: 2015*. Pew Research Center. Washington, DC: PewResearch Center; 2015. URL: <http://www.pewinternet.org/2015/10/29/technology-device-ownership-2015/> [accessed 2016-03-11] [WebCite Cache ID [6fwQOJ4As](https://www.webcitation.org/6fwQOJ4As)]
16. Catalyst. *With growth comes change: the evolving mobile landscape in 2015*. Toronto, ON: Catalyst; 2015. URL: <http://catalyst.ca/2015-canadian-smartphone-market/> [WebCite Cache ID [6fwQzn7fd](https://www.webcitation.org/6fwQzn7fd)]
17. Insights West. *Smartphone insights*. 2013 Jul. URL: [http://www.insightswest.com/wp-content/uploads/2013/12/BCAMASEP\\_InsightsWest\\_iamota\\_SmartphonesForIamota\\_17Dec2012\\_updatedPUBLICVERSION\\_DM.pdf](http://www.insightswest.com/wp-content/uploads/2013/12/BCAMASEP_InsightsWest_iamota_SmartphonesForIamota_17Dec2012_updatedPUBLICVERSION_DM.pdf) [accessed 2016-05-28] [WebCite Cache ID [6hqJQRZdu](https://www.webcitation.org/6hqJQRZdu)]
18. Pew Research Center. *Digital differences*. 2012 Apr. URL: [http://www.pewinternet.org/files/old-media/Files/Reports/2012/PIP\\_Digital\\_differences\\_041312.pdf](http://www.pewinternet.org/files/old-media/Files/Reports/2012/PIP_Digital_differences_041312.pdf) [accessed 2016-03-12] [WebCite Cache ID [6fwRNNQmk](https://www.webcitation.org/6fwRNNQmk)]
19. Dennison L, Morrison L, Conway G, Yardley L. Opportunities and challenges for smartphone applications in supporting health behavior change: qualitative study. *J Med Internet Res* 2013;15(4):e86 [FREE Full text] [doi: [10.2196/jmir.2583](https://doi.org/10.2196/jmir.2583)] [Medline: [23598614](https://pubmed.ncbi.nlm.nih.gov/23598614/)]
20. Ghorai K, Akter S, Khatun F, Ray P. mHealth for smoking cessation programs: a systematic review. *J Pers Med* 2014;4(3):412-423 [FREE Full text] [doi: [10.3390/jpm4030412](https://doi.org/10.3390/jpm4030412)] [Medline: [25563359](https://pubmed.ncbi.nlm.nih.gov/25563359/)]
21. McClure JB, Hartzler AL, Catz SL. Design considerations for smoking cessation apps: feedback from nicotine dependence treatment providers and smokers. *JMIR Mhealth Uhealth* 2016;4(1):e17 [FREE Full text] [doi: [10.2196/mhealth.5181](https://doi.org/10.2196/mhealth.5181)] [Medline: [26872940](https://pubmed.ncbi.nlm.nih.gov/26872940/)]
22. Bricker JB, Mull KE, Kientz JA, Vilardaga R, Mercer LD, Akioka KJ, et al. Randomized, controlled pilot trial of a smartphone app for smoking cessation using acceptance and commitment therapy. *Drug Alcohol Depend* 2014 Oct 1;143:87-94. [doi: [10.1016/j.drugalcdep.2014.07.006](https://doi.org/10.1016/j.drugalcdep.2014.07.006)] [Medline: [25085225](https://pubmed.ncbi.nlm.nih.gov/25085225/)]
23. Buller DB, Borland R, Bettinghaus EP, Shane JH, Zimmerman DE. Randomized trial of a smartphone mobile application compared to text messaging to support smoking cessation. *Telemed J E Health* 2014 Mar;20(3):206-214 [FREE Full text] [doi: [10.1089/tmj.2013.0169](https://doi.org/10.1089/tmj.2013.0169)] [Medline: [24350804](https://pubmed.ncbi.nlm.nih.gov/24350804/)]
24. Ubhi HK, Michie S, Kotz D, Wong WC, West R. A mobile app to aid smoking cessation: preliminary evaluation of SmokeFree28. *J Med Internet Res* 2015;17(1):e17 [FREE Full text] [doi: [10.2196/jmir.3479](https://doi.org/10.2196/jmir.3479)] [Medline: [25596170](https://pubmed.ncbi.nlm.nih.gov/25596170/)]
25. Baskerville NB, Struik LL, Hammond D, Guindon GE, Norman CD, Whittaker R, et al. Effect of a mobile phone intervention on quitting smoking in a young adult population of smokers: randomized controlled trial study protocol. *JMIR Res Protoc* 2015;4(1):e10 [FREE Full text] [doi: [10.2196/resprot.3823](https://doi.org/10.2196/resprot.3823)] [Medline: [25599695](https://pubmed.ncbi.nlm.nih.gov/25599695/)]
26. BinDhim NF, McGeechan K, Trevena L. Assessing the effect of an interactive decision-aid smartphone smoking cessation application (app) on quit rates: a double-blind automated randomised control trial protocol. *BMJ Open* 2014;4(7):e005371 [FREE Full text] [doi: [10.1136/bmjopen-2014-005371](https://doi.org/10.1136/bmjopen-2014-005371)] [Medline: [25037644](https://pubmed.ncbi.nlm.nih.gov/25037644/)]
27. Garrison KA, Pal P, Rojiani R, Dallery J, O'Malley SS, Brewer JA. A randomized controlled trial of smartphone-based mindfulness training for smoking cessation: a study protocol. *BMC Psychiatry* 2015;15:83 [FREE Full text] [doi: [10.1186/s12888-015-0468-z](https://doi.org/10.1186/s12888-015-0468-z)] [Medline: [25884648](https://pubmed.ncbi.nlm.nih.gov/25884648/)]
28. Hassandra M, Lintunen T, Kettunen T, Vanhala M, Toivonen H, Kinnunen K, et al. Effectiveness of a mobile phone app for adults that uses physical activity as a tool to manage cigarette craving after smoking cessation: a study protocol for a randomized controlled trial. *JMIR Res Protoc* 2015;4(4):e125 [FREE Full text] [doi: [10.2196/resprot.4600](https://doi.org/10.2196/resprot.4600)] [Medline: [26494256](https://pubmed.ncbi.nlm.nih.gov/26494256/)]
29. Valdivieso-López E, Flores-Mateo G, Molina-Gómez JD, Rey-Reñones C, Barrera Uriarte ML, Duch J, et al. Efficacy of a mobile application for smoking cessation in young people: study protocol for a clustered, randomized trial. *BMC Public Health* 2013;13:704 [FREE Full text] [doi: [10.1186/1471-2458-13-704](https://doi.org/10.1186/1471-2458-13-704)] [Medline: [23915067](https://pubmed.ncbi.nlm.nih.gov/23915067/)]
30. Fiore M, Jaen C, Baker T, Bailey W, Benowitz N, Curry S, et al. *Treating Tobacco Use and Dependence: 2008 Update*. Rockville, MD: US Department of Health and Human Services, Public Health Service; 2008.
31. Abrams LC, Lee WJ, Bontemps-Jones J, Ramani R, Mellerson J. A content analysis of popular smartphone apps for smoking cessation. *Am J Prev Med* 2013 Dec;45(6):732-736 [FREE Full text] [doi: [10.1016/j.amepre.2013.07.008](https://doi.org/10.1016/j.amepre.2013.07.008)] [Medline: [24237915](https://pubmed.ncbi.nlm.nih.gov/24237915/)]

32. Ubhi HK, Kotz D, Michie S, van Schayck OCP, Sheard D, Selladurai A, et al. Comparative analysis of smoking cessation smartphone applications available in 2012 versus 2014. *Addict Behav* 2016 Jul;58:175-181 [FREE Full text] [doi: [10.1016/j.addbeh.2016.02.026](https://doi.org/10.1016/j.addbeh.2016.02.026)] [Medline: [26950256](https://pubmed.ncbi.nlm.nih.gov/26950256/)]
33. Hoepfner BB, Hoepfner SS, Seaboyer L, Schick MR, Wu GW, Bergman BG, et al. How smart are smartphone apps for smoking cessation? A content analysis. *Nicotine Tob Res* 2015 Jun 4;18(5):1025-1031. [doi: [10.1093/ntr/ntv117](https://doi.org/10.1093/ntr/ntv117)] [Medline: [26045249](https://pubmed.ncbi.nlm.nih.gov/26045249/)]
34. Ritchie J, Spencer L. Qualitative data analysis for applied policy research. In: Bryman A, Burgess B, editors. *Analyzing Qualitative Data*. London: Routledge; 1994. 173-194.
35. Tong A, Sainsbury P, Craig J. Consolidated criteria for reporting qualitative research (COREQ): a 32-item checklist for interviews and focus groups. *Int J Qual Health Care* 2007 Dec;19(6):349-357 [FREE Full text] [doi: [10.1093/intqhc/mzm042](https://doi.org/10.1093/intqhc/mzm042)] [Medline: [17872937](https://pubmed.ncbi.nlm.nih.gov/17872937/)]
36. Fogg BJ. A behavior model for persuasive design. 2009. URL: [http://www.bjfogg.com/fbm\\_files/page4\\_1.pdf](http://www.bjfogg.com/fbm_files/page4_1.pdf) [accessed 2016-11-08] [WebCite Cache ID 6lrv71VVE]
37. van Gemert-Pijnen JE, Nijland N, van Limburg M, Ossebaard HC, Kelders SM, Eysenbach G, et al. A holistic framework to improve the uptake and impact of eHealth technologies. *J Med Internet Res* 2011;13(4):e111 [FREE Full text] [doi: [10.2196/jmir.1672](https://doi.org/10.2196/jmir.1672)] [Medline: [22155738](https://pubmed.ncbi.nlm.nih.gov/22155738/)]
38. Bandura A. Health promotion by social cognitive means. *Health Educ Behav* 2004 Apr;31(2):143-164. [doi: [10.1177/1090198104263660](https://doi.org/10.1177/1090198104263660)] [Medline: [15090118](https://pubmed.ncbi.nlm.nih.gov/15090118/)]
39. BinDhim NF, McGeechan K, Trevena L. Who uses smoking cessation apps? A feasibility study across three countries via smartphones. *JMIR Mhealth Uhealth* 2014;2(1):e4 [FREE Full text] [doi: [10.2196/mhealth.2841](https://doi.org/10.2196/mhealth.2841)] [Medline: [25098439](https://pubmed.ncbi.nlm.nih.gov/25098439/)]
40. Sayette MA, Martin CS, Hull JG, Wertz JM, Perrott MA. Effects of nicotine deprivation on craving response covariation in smokers. *J Abnorm Psychol* 2003 Feb;112(1):110-118 [FREE Full text] [Medline: [12653419](https://pubmed.ncbi.nlm.nih.gov/12653419/)]
41. Kreuter MW, Strecher VJ, Glassman B. One size does not fit all: the case for tailoring print materials. *Ann Behav Med* 1999;21(4):276-283. [Medline: [10721433](https://pubmed.ncbi.nlm.nih.gov/10721433/)]
42. DiClemente R, Salazar L, Crosby R. *Health Behavior Theory for Public Health*. Burlington, MA: Jones & Bartlett Publishers; 2013.
43. Clinical Practice Guideline Treating Tobacco Use and Dependence 2008 Update Panel, Liaisons, and Staff. A clinical practice guideline for treating tobacco use and dependence: 2008 update. A U.S. Public Health Service report. *Am J Prev Med* 2008 Aug;35(2):158-176 [FREE Full text] [doi: [10.1016/j.amepre.2008.04.009](https://doi.org/10.1016/j.amepre.2008.04.009)] [Medline: [18617085](https://pubmed.ncbi.nlm.nih.gov/18617085/)]
44. Struik LL, Baskerville NB. The role of Facebook in Crush the Crave, a mobile- and social media-based smoking cessation intervention: qualitative framework analysis of posts. *J Med Internet Res* 2014;16(7):e170 [FREE Full text] [doi: [10.2196/jmir.3189](https://doi.org/10.2196/jmir.3189)] [Medline: [25016998](https://pubmed.ncbi.nlm.nih.gov/25016998/)]
45. Ghorai K, Jahan S, Ray P, Chylinski M. mHealth for behaviour change: role of a smartphone based multi-intervention service for hypertension and diabetes in Bangladesh. *Int J Biomed Eng Technol* 2014;16(2):135-155. [doi: [10.1504/IJBET.2014.065655](https://doi.org/10.1504/IJBET.2014.065655)]
46. Zhu SH, Lee M, Zhuang YL, Gamst A, Wolfson T. Interventions to increase smoking cessation at the population level: how much progress has been made in the last two decades? *Tob Control* 2012 Mar;21(2):110-118 [FREE Full text] [doi: [10.1136/tobaccocontrol-2011-050371](https://doi.org/10.1136/tobaccocontrol-2011-050371)] [Medline: [22345233](https://pubmed.ncbi.nlm.nih.gov/22345233/)]

## Abbreviations

**FG:** focus group

**LGBTQ+:** lesbian, gay, bisexual, trans, queer, and other sexual minorities

**mHealth:** mobile health

**YYA:** youth and young adults

*Edited by P Sullivan; submitted 10.06.16; peer-reviewed by M Hassandra, P Ray, K DeLaughter; comments to author 02.10.16; revised version received 27.10.16; accepted 29.10.16; published 18.11.16.*

*Please cite as:*

Baskerville NB, Dash D, Wong K, Shuh A, Abramowicz A

*Perceptions Toward a Smoking Cessation App Targeting LGBTQ+ Youth and Young Adults: A Qualitative Framework Analysis of Focus Groups*

*JMIR Public Health Surveill* 2016;2(2):e165

URL: <http://publichealth.jmir.org/2016/2/e165/>

doi: [10.2196/publichealth.6188](https://doi.org/10.2196/publichealth.6188)

PMID: [27864164](https://pubmed.ncbi.nlm.nih.gov/27864164/)

©N Bruce Baskerville, Darly Dash, Katy Wong, Alanna Shuh, Aneta Abramowicz. Originally published in JMIR Public Health and Surveillance (<http://publichealth.jmir.org>), 18.11.2016. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Public Health and Surveillance, is properly cited. The complete bibliographic information, a link to the original publication on <http://publichealth.jmir.org>, as well as this copyright and license information must be included.

Original Paper

# Reaction on Twitter to a Cluster of Perinatal Deaths: A Mixed Method Study

Sarah Meaney<sup>1,2\*</sup>, MA; Leanne Cussen<sup>2\*</sup>, BSc (Med); Richard A Greene<sup>1\*</sup>, BAgrSc, MB, BCh, BAO; Keelin O'Donoghue<sup>2\*</sup>, PhD, MB, BAO, FRCOG, BCh

<sup>1</sup>National Perinatal Epidemiology Centre, Obstetrics and Gynaecology, University College Cork, Cork, Ireland

<sup>2</sup>Pregnancy Loss Research Group, Department of Obstetrics and Gynaecology, University College Cork, Cork, Ireland

\* all authors contributed equally

**Corresponding Author:**

Sarah Meaney, MA  
National Perinatal Epidemiology Centre  
Obstetrics and Gynaecology  
University College Cork  
Cork University Maternity Hospital  
Cork,  
Ireland  
Phone: 353 214205054  
Fax: 353 214205025  
Email: [s.meaney@ucc.ie](mailto:s.meaney@ucc.ie)

## Abstract

**Background:** Participation in social networking sites is commonplace and the micro-blogging site Twitter can be considered a platform for the rapid broadcasting of news stories.

**Objective:** The aim of this study was to explore the Twitter status updates and subsequent responses relating to a number of perinatal deaths which occurred in a small maternity unit in Ireland.

**Methods:** An analysis of Twitter status updates, over a two month period from January to March 2014, was undertaken to identify the key themes arising in relation to the perinatal deaths.

**Results:** Our search identified 3577 tweets relating to the reported perinatal deaths. At the height of the controversy, Twitter updates generated skepticism in relation to the management of not only of the unit in question, which was branded as unsafe, but also the governance of the entire Irish maternity service. Themes of concern and uncertainty arose whereby the professional motives of the obstetric community and staffing levels in the maternity services were called into question.

**Conclusions:** Twitter activity provides a useful insight into attitudes towards health-related events. The role of the media in influencing opinion is well-documented and this study underscores the challenges that clinicians face in light of an obstetric media scandal. Further study to identify how the obstetric community could develop tools to utilize Twitter to disseminate valid health information could be beneficial.

(*JMIR Public Health Surveill* 2016;2(2):e36) doi:[10.2196/publichealth.5333](https://doi.org/10.2196/publichealth.5333)

**KEYWORDS**

social media; health care services; maternity; perinatal death; Twitter; infodemiology; infoveillance

## Introduction

During pregnancy women are invested in seeking out a considerable amount of information in relation to pregnancy and the services that are available to them [1]. To date the provision of pregnancy-related information during the antenatal period has been through more traditional media [1]. These, including leaflets, magazines and advertisements, channel the

information directly from source to target audience [2]. Thus, women are limited to the passive viewing of pregnancy-related content, which has been created for them [3].

Rapid development in Web-based technologies has seen a shift in how women now access pregnancy-related information. A recent study has shown that 95% of pregnant women in Ireland use the Internet for pregnancy information [1]. The transition from more traditional to digitally-based media may be related

to two issues: (1), women of child-bearing age in developed countries have access to a wide array of technologies including personal computers, laptops, tablets and smartphones; (2) women may have more confidence in the information that they receive online given how Stapleton et al found that traditional materials such as leaflets or books were considered not only limited and biased but that the information provided was considered dated [4].

Health care providers and policymakers need to address the information needs of pregnant women [5]. The benefit of Internet-based technologies for women is that they are no longer passive users. The development of these technologies ensures that information provision is consumer-centered whereby not only users are encouraged to interact with others but also to create and share content through multiple digital channels [6]. Bernhardt et al outline how this revolution in communication has already affected health care as a new generation of e-patients has emerged [3]. Digital media has empowered these health care consumers, allowing them to be even more engaged in their care and in turn to influence their current and future service provision [3].

The expansion of social network platforms in particular has driven these developments. Participation in social networking websites such as the micro-blogging site Twitter is now commonplace. It is reported that 38 million adults in the United Kingdom access the Internet daily, with over half of the population participating in social networking [7]. A national cross-sectional survey from 2015 indicated that 26% of the Irish population use Twitter, with one in three of those people using Twitter daily [8]. However, a tweet extends far beyond the individual status as it includes its audience; those who may read the tweet, retweet and/or reply [9]. As any status update can be seen quickly by a very large audience [10], Twitter must be considered a platform for rapid and immediate communication. These sites allow individuals to share their thoughts on the information they are currently consuming [10]. Such rapid communications among these new e-patients are very likely to be influential in how people consume information in relation to health care services [10]. Health care consumers now have a new digital space where they can discuss information that is provided to them and evaluate their health care services as they experience it, either while attending a clinic or while watching a TV broadcast [10]. This phenomenon has already been observed in the United Kingdom where reforms in the National Health Service have been influenced by social commentary on Twitter [11].

Social media is transforming health-related research [12]. Studies have been undertaken from a wide range of disciplines, from epidemiology where real time responses to pandemics are analyzed [13] to the behavioral sciences where the way patients consume online health-related information is examined [3,14]. The manner in which women both access and respond to pregnancy-related information needs to be evaluated as it is likely to be influential in relation to women's decision making [4].

Reports in the media related to pregnancy and birth are common and these reports are often emotive [15]. Such reporting may

have a detrimental impact as it can potentially misinform pregnant women and may possibly result in confusion and anxiety [15]. How news is communicated has altered dramatically as websites, social media, and 24-hour rolling broadcasts have seen rapid growth and increased popularity [15]. Since 2012, there has been considerable national and international media coverage reporting a number of adverse incidents within the Irish maternity services, which in turn have resulted in a number of independent enquiries into the services. The media coverage of these adverse incidents stimulated much debate, including discussion in relation to a cluster of perinatal deaths in Midland Regional Hospital, Portlaoise in 2014. Although there is ongoing research on the causes of perinatal death, which aims to reduce its prevalence, there is still persistent stigma associated with perinatal death [16]. Stillbirth, in particular, is often referred to as a silent loss [16]. Thus, our study aimed to explore the reaction on Twitter to the perinatal deaths in order to gain insight into the understanding and perception of perinatal death in the Republic of Ireland.

## Methods

### Setting

In 2014, the Irish Central Statistics Office estimated that the Irish population was 4,593,100 [17]. In Ireland, the Maternity and Infant Care Scheme grants women ordinarily resident in Ireland access to free maternity services. This public service is provided by both a general practitioner and a maternity health care provider. The majority of births occur in one of the 19 obstetric led units in Ireland [18]. Ireland has the highest birth rate in Europe with 15.6 births per 1000 population [19]. In 2014, there were 67,462 births of which 330 were stillbirths; defined here as an infant born with no sign of life weighing 500 grams or more and/or having a gestational age of 24 weeks or more [20].

Raidió Teilifís Éireann (RTE) is an Irish television broadcaster. As a national public-service media organization the service which it provides is free to air [21]. RTE's flagship current affairs program is *Primetime*. In 2015 when this study was conducted, RTE had 205,000 followers on Twitter and the *Primetime* account had 69,000 followers. On January 30, 2014 RTE aired a program entitled "Fatal Failures" on *Primetime*. The program was concerned with a cluster of perinatal deaths that occurred in a maternity hospital located in the Midlands in the Republic of Ireland. Following this program the Health Minister, James Reilly, requested the Chief Medical Officer to prepare a report on the issues identified in the program. The Chief Medical Officer published the report on February 28, 2014 [22]. The Health Minister also requested the Health Information Quality Authority, an independent authority who is responsible for assessing quality and safety of health care services, to undertake a review of the hospital in question.

### Design

As this study was an observational study of Twitter status updates in relation to the reported perinatal death, a mixed methods approach was adopted. By utilizing a mixed methods design the study was able to benefit from analyzing the data both quantitatively and qualitatively. For the purpose of this



study the data were initially quantitatively analyzed in order to assess the frequency of status updates, the demographic profiles of users, and to ascertain the potential reach of the status updates. Secondly, qualitative analysis was employed to generate themes from the content that Twitter users shared publically.

### Search Strategy

Status updates in English were manually searched using the Twitter search function on its website. A search was undertaken on all public status updates from January 29, 2014 to March 31, 2014 relating to the perinatal deaths in the Republic of Ireland. The two month period was chosen as research indicates that public interest spike around the time of the event and decline rapidly thereafter [23]. This time frame was chosen to allow for the examination of the immediate response to the reporting of the perinatal deaths and the subsequent reports published in relation to the perinatal deaths.

Seven searches were conducted independently by two researchers (SM and LC) which included: “fatal failures” (the name of the episode regarding the perinatal deaths), “rtept” (the current affairs program which aired the episode), “death of a baby”, “maternity”, “stillbirth”, “perinatal death”, and finally we searched the name of the hospital where the deaths occurred. Initially the searches were limited to using hashtags (#) which are useful to search for content on Twitter as they group messages on a specific event together. However, we found that this limited the search. Therefore, we chose to proceed using the terms alone for the search process as it produced more results, including the hashtags related to the perinatal deaths. All data which were extracted from each of the searches were stored in Microsoft Excel. Each tweet was then reviewed and assessed for inclusion in the study.

Once these searches were complete, any demographic information available was collected from the users’ public profiles. This included their biography, their location, and the number of people who were following the user at the time of data collection in 2015.

Only data which were publically available were collected and no attempts were made to contact any individual; therefore, no ethical approval was sought for this study. Despite these data being available to public, there is still an onus to ensure that ethical standards are met. Therefore, in line with other similar studies [14,24,25] personal identity information, including individuals’ Twitter usernames, have been removed from the example tweets presented below.

### Analysis

Both quantitative and qualitative methods were utilized for this study. First, in order to determine the volume of social media communication in relation to the perinatal deaths, descriptive statistics of all tweets, retweets and replies were calculated. If a user’s biography was available, this was coded by the researchers (SM and LC) in order to categorize the demographic profiles of the users. The biography on Twitter is limited to 160 characters and is the user’s self-description. Therefore, the demographic data reported here is self-identified by the user.

Consequently, users were broadly grouped into the following categories which are reported here: parent, media outlet, media personnel, politics, and health. These categories were not considered mutually exclusive whereby, for example, a user may describe themselves as a “midwife and mother of two children” and therefore, would be considered both a parent and a health care professional. In order to determine the potential reach of the status updates, the number of people that were following the user was collected.

Given that this is an observational study, a qualitative methodology which is more descriptive rather than interpretative was chosen. Consequently, a thematic analysis of the text within the Twitter status updates was then performed electronically using Nvivo 10 software (QSR International Pty Ltd, Doncaster, Australia). The analytic process, as outlined by Braun and Clarke (2006), involves familiarization with the data whereby the researchers read and re-read each tweet which was then coded individually [26]. These preliminary codes were reviewed and similar individual codes were identified and grouped together as categories. The final themes were then agreed by grouping related categories together.

## Results

### Quantitative Results

Over the two month period from January 29, 2014 to March 31, 2014, 3577 Twitter status updates from 1276 profiles relating to the perinatal deaths in Midland Regional Hospital, Portlaoise were identified. Of these status updates 45.15% (1615/3577) were tweets, 38.92% (1392/3577) were retweets, and 15.94% (570/3577) were replies.

As illustrated in [Figure 1](#), around 39.84% (1425/3577) of status updates were posted between the 29<sup>th</sup> and 31<sup>st</sup> of January coinciding the airing of the current affairs program which investigated the perinatal deaths. Almost half of all status updates which were replies were posted on the 30<sup>th</sup> of January (48.4%; 276/570). The second largest peak of status updates (21.5%; 770/3577) related to the publication of the findings from the Chief Medical Officer’s report.

The individual profiles of those who posted status updates were analyzed to discern demographic characteristics. Of the 1276 profiles, a biography was available for 1139 (89.3%). Profiles indicated that those who self-identified as being involved in media, health care and/or politics accounted for almost two thirds of status updates (62.87%; 2249/3577). More than 1 in 10 status updates were by those who self-identified as a parent (11.85%; 424/3577). [Table 1](#) outlines the distribution of status updates among these groups. Profiles which identified the user as either a media outlet or personnel working for a media outlet accounted for over one third (36.82%; 1317/3577) of status updates. Of note, media outlets were more likely to create content; whereby the majority (83.3%) of their updates were tweets compared to other users. Parents and those involved in health care were more likely to question or discuss content with 29.0% and 24.2% of their updates being replies.

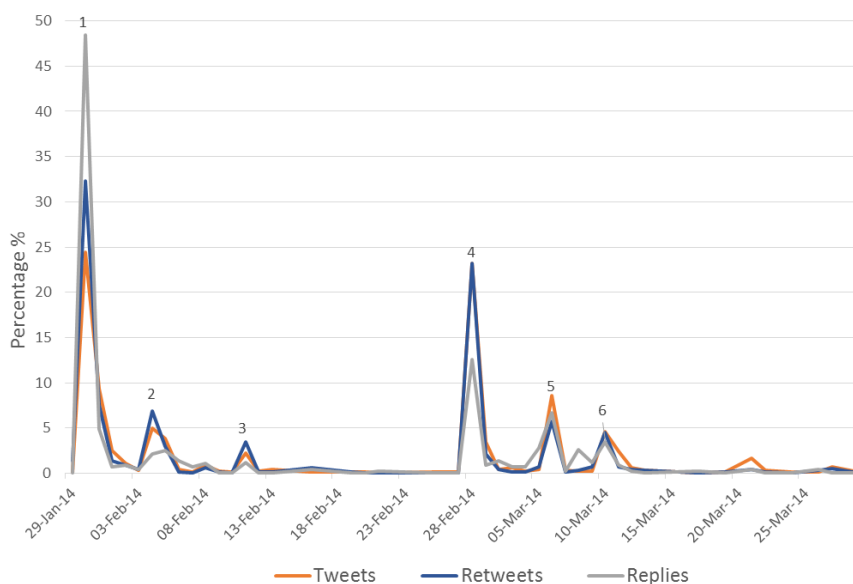
**Table 1.** Sample characteristics.

	All status updates (n=3577)	Media outlets (n=540)	Media personnel (n=777)	Health care profes- sional (n=467)	Politics (n=465)	Parent (n=424)
Tweet n (%)	1615 (45.20)	450 (83.3)	364 (46.8)	178 (38.1)	179 (38.55)	133 (31.4)
Retweet n (%)	1392 (38.92)	90 (16.7)	310 (39.9)	176 (37.7)	202 (43.4)	168 (39.6)
Reply n (%)	570 (15.94)	0 (0)	103 (13.3)	113 (24.2)	84 (18.1)	123 (29.0)
Mean number of followers	11,709	59,437	5667	1902	3663	1138

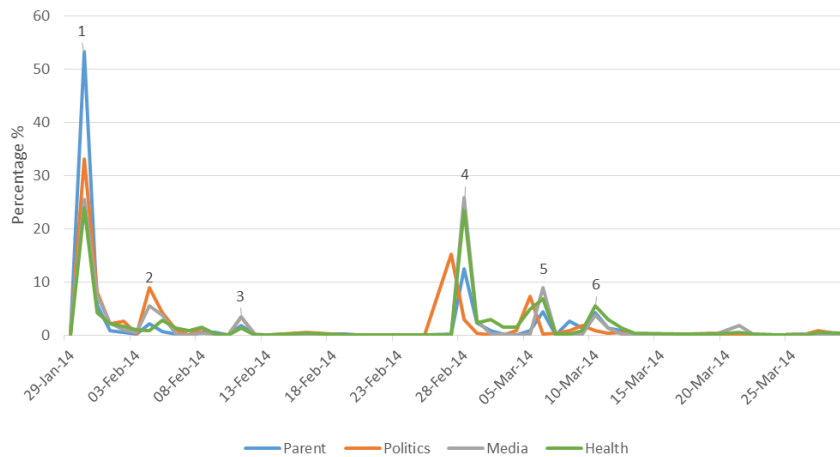
Of the profiles related to health care (n=111), almost one third (29.7%) were from a diverse range of support organizations and online health care businesses, such as those who provide health, fitness, and nutrition services online. One in five identified as either a midwife or a nurse. Fourteen percent identified as a medical doctor or consultant; however, none were from the field of obstetrics and gynaecology. Of the political profiles (n=122), one quarter identified themselves as a political figure in Ireland such as a Member of the Irish Parliament or a local Councillor. One in six profiles saw individuals identify themselves as activists (16.4%).

Over half of the status updates (53.3%; 226/424) posted by parents occurred on January 30<sup>th</sup>(Figure 2). One of the largest volumes of status updates, 13.2% (62/465), by those involved in politics occurred when the Minister for Health made an emotional statement while announcing a hospital investigation was to be undertaken. Almost one quarter (23.6%; 110/465) of status updates by those involved in health care occurred when the findings from the Chief Medical Officer’s report were published.

**Figure 1.** Frequency of status updates by date: (1) Current affairs television program Primetime air the episode entitled Fatal Failures relating to a cluster of perinatal deaths in Portlaoise hospital in the Republic of Ireland; (2) Minister for Health makes an emotional statement after meeting the families who had a perinatal death and announces the investigation by the Chief Medical Officer; (3) A mother releases a statement that she was only made aware that an investigation was undertaken on the perinatal death as a result of the Primetime program; (4) Findings from the Chief Medical Officer’s report are published; (5) It is confirmed that an independent investigation of the hospital will be undertaken by the Health Information Quality Authority; (6)The Health Service Executive confirm that a perinatal death occurred in Portlaoise hospital on March 8, 2014.



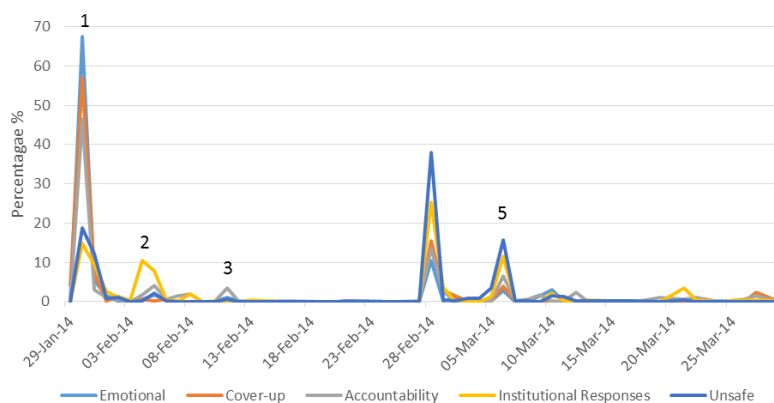
**Figure 2.** Frequency of status updates by date and user: (1) Current affairs television program Primetime air the episode entitled Fatal Failures relating to a cluster of perinatal deaths in Portlaoise hospital in the Republic of Ireland; (2) Minster for Health makes an emotional statement after meeting the families who had a perinatal death and announces the investigation by the Chief Medical Officer; (3) A mother releases a statement that she was only made aware that an investigation was undertaken on the perinatal death as a result of the Primetime program; (4) Findings from the Chief Medical Officer’s report are published; (5) It is confirmed that an independent investigation of the hospital will be undertaken by the Health Information Quality Authority; (6) The Health Service Executive confirm that a perinatal death occurred in Portlaoise hospital on March 8, 2014.



**Qualitative Results**

**Themes**

**Figure 3.** Frequency of status updates by date and theme: (1) Current affairs television program Primetime air the episode entitled Fatal Failures relating to a cluster of perinatal deaths in Portlaoise hospital in the Republic of Ireland; (2) Minster for Health makes an emotional statement after meeting the families who had a perinatal death and announces the investigation by the Chief Medical Officer; (3) A mother releases a statement that she was only made aware that an investigation was undertaken on the perinatal death as a result of the Primetime program; (4) Findings from the Chief Medical Officer’s report are published; (5) It is confirmed that an independent investigation of the hospital will be undertaken by the Health Information Quality Authority; (6) The Health Service Executive confirm that a perinatal death occurred in Portlaoise hospital on March 8, 2014.



Qualitative analysis of the tweets resulted in the identification of five key themes: emotional reactions, cover-ups, accountability and governance, institutional responses, and unsafe maternity services (Figure 3).

**Emotional Reactions**

The majority of tweets, including retweets and replies, in this theme were characterized as negative responses where anger, distress, and upset were communicated. As the current affairs

program “Fatal Failures” aired, tweets indicated that the viewers of the episode were shocked and upset by what was reported. A number of status updates indicated the frightening and distressing nature of the program (Textbox 1).

Tweeters sympathized with the families for the perinatal loss they had experienced. The tweets reveal how individuals expressed empathy for these parents as they tried to comprehend

the loss experienced, thankful that they themselves had not experienced such tragedy ([Textbox 1](#)).

**Textbox 1.** Examples of tweets showing emotional reactions.

“Frightening & Distressing”

*That was so incredibly sad #rtept. My son, nieces & nephews were born in Portlaoise & I'll go to bed tonight counting my blessings.*

*Heartbreaking and chilling at the same time- has unsettled me #expectingno3 #rtept*

*So terrible looking at these beautiful babies that never got a chance to live their lives. Very distressing. #rtept*

“Empathy for Parents”

*This is so upsetting. Feeling blessed that my little baby was delivered safely in Portlaoise. My heart goes out to those not as lucky #rtept*

*Absolutely shocking #rtept Report Fatal Failures, heart goes out to families, should never have happened & should never happen again!*

*Just bawled my eyes out all the way through @RTE\_PrimeTime My heart goes out to those families who lost their babies unnecessarily #rtept*

*My heart goes out to the mothers and fathers featured on #rtept #primetime I don't think I'd have the strength if I was in their shoes*

*@RTE\_PrimeTime fantastic report, impossible viewing, inspiring families, cruel cruel system. Congrats. Hard not to be angry.*

“Shock & Anger”

*Can only imagine the torture of the uncertainty, questions, what ifs, maybes, if onlys these parents must have gone through for years #rtept*

*@RTE\_PrimeTime absolutely shocking behavior. Hard to watch.*

*Cannot believe what they did at Portlaoise hospital, an utter disgrace to the medical profession. #Sickening*

*Heartbreaking watching #rtept Not good enough #HSE Shame on the hospital. My thoughts are with those brave families. Devastated lives.*

*Disgusted by the HSE and Portlaoise Hospital. Incompetent callous and cowardly #rtept*

The shock and anger expressed on Twitter were aggravated by the portrayal of the hospital staff and management. The fundamental values of care and compassion, which are normally attributed to those who provide health-related services, were at odds with the televised representation of those who were responsible for the care of the families who had experienced perinatal loss ([Textbox 1](#)).

### Cover-ups

Tweeters indicated that they believed the hospital management's priority was to cover up the events surrounding the perinatal deaths rather than focusing on the appropriate care for the patients ([Textbox 2](#)).

**Textbox 2.** Examples of tweets showing cover-ups and frustration with both the Irish Government and the Health Service Executive.

“Cover-ups”

*#rtept on neonatal deaths horrific viewing #HSE comes out poorly obstructing info to families, spokesman dodging issues, evasive.*

*#rtept I find the cover-up so much more distressing than the actual deaths*

*It was appalling for all but the mother who was left believing that something she did may have caused the death was just cruel*

*How do they sleep at night knowing they've sent somebody home not knowing why they're really leaving hospital without their baby. #rtept*

“Frustration”

*Are we to assume reports were made available only because #rtept were investigating?*

*Reilly seeks report on baby deaths...surely our Health Minister doesn't need a TV docu to prompt an investigation*

*#rtept shows again & again how reports are published & never implemented in our hospitals so women & babies die*

*We never learn! Barely a word about the HIQA #Savita report recommendations. Where is the sense of urgency about improving standards. #rtept*

*HSE guy - I regret ""IF"" any actions..... The usual PR speak. Maddening #rtept*

*Independent Hiqa inquiry into baby deaths in #Portlaoise hospital is essential. No point in HSE investigating itself. Families deserve truth*

The Irish national broadcaster, and the journalists involved in the investigation, were praised for their role in informing the public firstly of the occurrence of the perinatal deaths but also of the suppression of information by the hospital.

*#rtept Pubic Service Broadcasting at its best tonight! Harrowing stories of avoidable infant deaths at Portlaoise maternity hospital.*

*Common theme in Ireland on internal investigations, first instinct is to cover up, obstruct and frustrate. #rtept*

*Midwives in Portlaoise knew trouble was brewing and they were ignored. Maybe social media is the way forward so the public is informed.*

Tweets indicated frustration with both the Irish Government and the Health Service Executive, whereby their actions were seen as reactive and defensive (Textbox 2). It was suspected that the only motivating factor to investigate the health service

was as a result of the expose by the Irish broadcaster rather than genuine concern for ensuring that health care standards are met.

### **Accountability and Governance**

Tweets reveal immense dissatisfaction with the health care authorities as it was believed that if the recommendations of previous investigations on the Irish Maternity Services had been implemented these perinatal deaths would have been avoided (Textbox 3).

The tweets also revealed concern that there are no obvious implications when recommendations are not implemented or adhered to. Moreover, these tweets revealed how individuals appeared to be resigned to the fact that no official within Government or the health care authorities would be identified and held responsible for these failures (Textbox 3).

Given the belief that neither the Government or the health care authorities would take appropriate action, a criminal investigation was endorsed.

**Textbox 3.** Examples of tweets showing accountability and governance.

“Immense Dissatisfaction”

*#rtept sounds like complacency was rife in #Portlaoise no action taken on recommendations leading to unnecessary deaths, CTG, oxytocin*

*How many reviews and investigations do we need to have before change occurs - women and babies and all families deserve better #rtept*

“No Implication, No Responsibility”

*And tomorrow we Will discover that nobody Will be held accountable! #hse #rtept*

*The frustrating thing is nothing will change, where's the accountability, where is the governance. #Rtept yes but we hear that “the hospital” is to blame, convenient to blame a building #noaccountability*

*James Reilly YOU are the head of the health service and have failed people across the board, change must come from the top down #portlaoise*

*It's so important that someone is held responsible for the deaths of the babies in Portlaoise Hospital. Can't be allowed to fade from media.*

“Criminal Investigation”

*Isnt failure to act.....negligence, and when it results in death, manslaughter...? #rtept*

*@Newstalkfm what was allowed to continue is worse than negligence, it was criminal, the gardai should be sent into that hospital*

### **Institutional Responses**

Governmental responses, which were shared on Twitter by those in media, initially rallied behind the online reactions describing the events which occurred in Midland Regional Hospital, Portlaoise as inappropriate and unacceptable. The Government made assurances that a thorough investigation would be undertaken and the findings from this investigation would be acted upon to safeguard against similar events happening again (Textbox 4).

Twitter was not utilized as a platform by any health care authority to release a statement in relation to the perinatal deaths. The initial response from medical institutions and professionals from the obstetric community utilized traditional methods of communication to inform the public that the maternity services were safe, making reference to the countries' rates of perinatal death. These statements were then edited and tweeted by those working within media (Textbox 4).



**Textbox 4.** Examples of tweets showing government and institutional responses.

“Governmental Responses”

*Emotional Health Minister says Portlaoise concerns will be addressed*

*Health Minister James Reilly says he was deeply disturbed by the RTE Prime Time revelations about child deaths at Portlaoise Hospital*

*Kenny: No family should have to fight for truth in our health system (via @thejournal\_ie)*

*Minister Shatter calls the manner in which Portlaoise Hospital treated families, as revealed in recent days, inexcusable.*

“Institutional Responses”

*The @RCPI\_ObsGyn has issued a statement to reassure people about Irish maternity services following last night's #rtept programme.*

*Prof Fionnuala McAuliffe, @RCPI\_ObsGyn, says Ireland is a very safe country in which to have a baby, with low rates of perinatal deaths.*

As the public's concern about the Irish maternity services rose it became more evident that there may be possible implication on service attendance. The Government responded by focusing on reassuring the public that these services were indeed safe.

*EK: Portlaoise “will ensure that Ireland will continue to be recognised as one of the safest countries in the world in which to give birth”*

*Gilmore reassures women over maternity services*

### **Unsafe Maternity Services**

Concern and uncertainty arose whereby the professional motives of the obstetric community and the Government were called into question. Status updates over the two months indicated the skepticism that was generated in relation to the management of not only the unit in question, which was branded as unsafe, but also the governance of the entire maternity service in Ireland (Textbox 5).

**Textbox 5.** Examples of tweets showing skepticism and doubts surrounding safety.

“Skepticism”

*Optimum ratio of midwives to patients 1:28. Ratios at Portlaoise hospital 1:75. How can this be possible?#unsafelabours #portlaoise maternity service cannot be regarded as safe and sustainable within its current governance arrangement*

*I wonder if all maternity units were looked at would the others come out squeaky clean? So sad for those women and their families #rtept*

*The claim that Ireland is one of the safest countries in which to give birth ringing ever more hollow #rtept*

“Doubts Surrounding the Safety”

*HSE probe fifth baby death at Portlaoise Hospital as damning report brands it “unsafe”*

*I suspect you will be hearing a lot about 'never events' and Portlaoise Hospital in the coming days. CMO report must be read to be believed.*

*In 2006, 08, 09 & 12, there were 4 neonatal deaths at the maternity unit in Portlaoise from 'never events' according to CMO*

*Portlaoise report recommends adverse events in low risk pregnancies to be deemed “never events”.*

The publication of the investigation by the Chief Medical Officer further reinforced doubts about the safety of Portlaoise hospital. In the report the Chief Medical Officer states that, in Ireland in a low-risk pregnancy, any maternal or perinatal death associated with labor or delivery are to be documented as perinatal “never events”[15].

Tweets indicated that there was a lack of confidence in the quality of care currently being provided, with some women tweeting their reluctance to engage with or attend the services as they did not consider the services safe.

*I'd rather give birth on the side of the road than in that hospital #RtePT #HSE heartbreaking stuff*

*Heartbreaking and chilling at the same time- has unsettled me #expectingno3 #rtept*

*That won't ease my worries though, due to have baby in Portlaoise in early July :(*

*As someone due to give birth soon, I'll [be] watching that CTG trace monitor like a hawk after #rtept & demanding fast action if needed.*

## **Discussion**

### **Principal Findings**

From January 29 till March 31, 2014 there were 3577 status updates; including tweets, retweets and replies, posted on Twitter relating to a cluster of perinatal deaths that occurred in a maternity unit in the Republic of Ireland. Of these status updates, 40% were posted during January 29-31 which coincided with the airing of the current affairs program which brought the perinatal deaths to the attention of the public.

## Limitations

The content of social media can be exploited by health care authorities whereby an analysis of tweets allows health care authorities to identify and respond to concerns [3,13]. However, our study may be limited by a few factors. First, the aim of the study was to evaluate the response on Twitter to the perinatal deaths; however, we restricted our search to publicly available status updates. Twitter is an open forum where it is possible for connections to be nonreciprocal as a person may choose to follow an account and may not be followed in return. Yet, there are exceptions whereby account holders are given the opportunity to protect their account and make status updates available solely to those they give permission to. Twitter also provides the facility for users who are following each other to direct message each other, these messages are private and are not searchable through the search facility on Twitter. Second, our study may be limited by our search terms; however, we believe our search strategy was comprehensive as our list of terms was initially developed independently by two of the researchers (SM and LC) and that all terms identified by both researchers were included in the final list of search terms. The data from this study show that interest in the events spiked and declined quickly which is similar to previous published data; however, it would be of interest to examine the perception of perinatal death over a longer period of time to ascertain any differences in public reaction. Finally, studies have illustrated that Twitter is not fully representative of the general population [25] and therefore the results may be limited by selection bias. It would be of interest to determine if these findings would be observed following the examination of other social networking sites such as Facebook and/or following investigations which utilize more conventional social research methods offline.

## Comparison With Prior Work

The findings of this study support the statement, as reported by Ampofo et al, that people now use digital space to instantaneously evaluate and share their experiences of health care services while, in this instance, watching a national current affairs TV broadcast [10]. By analyzing this content our study identified a number of key themes highlighting the concern about the events which occurred in the hospital but also regarding the governance of the entire health care service. During the course of the television broadcast the tweets indicated that individuals were both shocked and distraught by the events which resulted in four perinatal deaths. Over the course of the broadcast the sentiment observed in the tweets transitioned from distress to anger. Of particular concern to the online community was the manner in which the hospital management were seen to have made attempts to suppress information from parents about the events surrounding their babies' deaths. In the following days, as more details emerged and were shared by the media about the cases, the governance of the maternity services and the health care service as a whole were called into question. Our study identified such a level of dissatisfaction with the governance that a demand for a criminal investigation was called for. Similar to Burnap 2014 [27], our findings illustrate that the frequency of tweet and retweets peaked in line with specific events; initially broadcasting of the television program and

subsequently the announcement of an enquiry and the publication of the enquiry.

Research in relation to health communication is now focused on the participatory nature of the Internet with particular reference to social media [13]. This reveals how the public can play a larger role in the various stages of knowledge translation which includes information generation, filtering, and as well as knowledge amplification [13]. Our findings highlight the participatory nature of social media, in particular the filtering and amplification of knowledge generated around the perinatal deaths. This study found that almost half of all status updates which were replies were posted on the 30th of January when the TV episode "fatal failures" was aired. Replies, when individuals were posting a response to a tweet, were indicative of individuals either supporting a statement or sentiment posted or querying the content which was posted. Chou et al state how this process has transformed the pattern of health-related communications, whereby online information sharing is considered more democratic given that it can be controlled by the patients, who share the information of importance to them [28].

However, these developments have raised concerns among health professionals and policy makers [29]. Due to the nature of social media information can be generated and circulated to a wide audience very quickly. Thus, although unintended, Chou et al state that negative health impacts due to misinformation can occur [28]. The findings from this study indicate that the deaths which occurred were perceived as preventable and that any future perinatal deaths should be prevented. Thus, given that the tweeters were becoming increasingly distrusting of the Health Service Executive and the Government, our data would suggest that the users considered the information provided online and through the media as more credible. This finding is in line with those of Coleman et al whereby people, in particular those from a lower socioeconomic status, believed that information posted online by those similar to them was more credible [30]. Peterson et al found that online users' perception of credibility varied and studies indicate that credibility of online information is linked not only with expertise but also trustworthiness [31]. This is illustrated in our study by the tweets in response to statements from both the obstetric community and the government. Efforts to reassure the women that the maternity services were safe, by making reference to perinatal statistics, were considered deceptive and were believed to be misleading. This perception was reinforced following the publication of the Chief Medical Officer's report which stated these deaths should from now on be considered as perinatal "never events" and that if such an event were to occur in the future, no reassurance can be derived from summary statistics such as perinatal mortality rates [22].

This study found that one third of all the content generated on Twitter in relation to the perinatal deaths were from media outlet accounts or media personnel. Almost all the content tweeted by media outlets were tweets of original content including news reports and updates. Media personnel were more likely to retweet the information generated by media outlets. This activity is suggestive of the influence the media has within digital spheres and its potential to influence not only perceptions but

also patients' decision making in relation to health care services. Ampofo et al refer to this process as "mediatisation" whereby the logic of the media guides behaviors and decision making throughout society [10]. This phenomenon was also identified in a study undertaken by Donelle and Booth who demonstrated that tweets and public discussions related to health promotion were shaped by a political-media social dynamic [24]. These authors concluded that it is important to determine how the influence of this dynamic on the representation of health, through social media, impacts on the public perception and interaction with health care [24]. Our findings suggest that during this period tweeters perceived the Irish maternity services as unsafe. A recent news article has indicated that there has been a 12% reduction in the number of births in the hospital under review, with the Health Service Executive confirming that in the direct aftermath of the controversy a reduction in attendances at booking clinics was observed [32].

The findings from this, and other similar studies, have shown that Twitter may have the potential to influence patients' decision making and behavior. Twitter was not utilized as a platform by any health care organization or authority to release a statement in relation to the perinatal deaths. One consequence of this is that the message was broadcasted through the media, which potentially inhibits the ability for the intended message to be delivered to its target audience. Lagan et al stress that

health professionals must acknowledge that decision making is influenced by the information which is sought and consumed by patients online [5]. These authors state that there is a need for health professionals to engage in this process and that this engagement would allow them to direct patients to both comprehensive and accurate information. Thackery et al further state that it is important that health care professionals and policy makers engage and exploit the participatory nature of these technological developments [33]. These new e-patients expect interaction and that social media not be used as "virtual pamphlet walls" [24].

## Conclusion

Twitter activity provides a useful insight into attitudes towards health-related events. The role of the media in influencing opinion is well documented and this study underscores the challenges that clinicians face in light of an obstetric media scandal. Given that patients are now likely to access health-related information online, it is imperative that health care providers are meeting the needs of potential service users. Our study highlights the need to exploit social media effectively in order for health care providers and policy makers to identify and respond to the concerns in relation to health care services. Further study to identify how the obstetric community could develop tools to utilize social media sites, such as Twitter, to disseminate valid health information could be beneficial.

## Conflicts of Interest

None declared.

## References

1. O'Higgins A, Murphy O, Egan A, Mullaney L, Sheehan S, Turner M. The use of digital media by women using the maternity services in a developed country. *Ir Med J* 2014;107(10):313-315. [Medline: [25556255](#)]
2. Lefebvre R. *Social Marketing Quarterly*. 2007. The new technology: the consumer as participant rather than target audience URL: [http://socialmarketing.blogs.com/Publications/SMQ-The\\_consumer\\_as\\_participant\\_2007.pdf](http://socialmarketing.blogs.com/Publications/SMQ-The_consumer_as_participant_2007.pdf) [accessed 2016-06-17] [WebCite Cache ID 6iKsIVLdH]
3. Bernhardt J, Mays D, Kreuter M. Dissemination 2.0: closing the gap between knowledge and practice with new media and marketing. *J Health Commun* 2011;16 Suppl 1:32-44. [doi: [10.1080/10810730.2011.593608](#)] [Medline: [21843094](#)]
4. Stapleton H, Kirkham M, Thomas G. Qualitative study of evidence based leaflets in maternity care. *BMJ* 2002 Mar 16;324(7338):639 [FREE Full text] [Medline: [11895821](#)]
5. Lagan B, Sinclair M, Kernohan W. What is the impact of the Internet on decision-making in pregnancy? A global study. *Birth* 2011 Dec;38(4):336-345. [doi: [10.1111/j.1523-536X.2011.00488.x](#)] [Medline: [22112334](#)]
6. O'REILLY T. What is Web 2.0: Design patterns and business models for the next generation of software. *Communications & Strategies*. *Communications & Strategies* 2007;1:17 [FREE Full text]
7. Office for National Statistics. ONS Research Bulletin. 2014. Internet Access - Households and Individuals, 2014 URL: <http://www.ons.gov.uk/peoplepopulationandcommunity/householdcharacteristics/homeinternetandsocialmediausage/bulletins/internetaccesshouseholdsandindividuals/2014-08-07> [accessed 2016-06-21] [WebCite Cache ID 6iQwqVBkF]
8. Ipsos M. Newsletter May. 2015. *Social Networking Quarterly* URL: <http://www.ipsosmrbi.com/social-networking-quarterly-survey-may-15.html> [accessed 2016-06-21] [WebCite Cache ID 6iQwyoitV]
9. Chadwick A. Britain's first live televised party leaders' debate: From the news cycle to the political information cycle. *Oxford Journals, Parliamentary Affairs* 2011;64(1):24-44.
10. Ampofo L, Anstead N, O'Loughlin B. TRUST, CONFIDENCE, AND CREDIBILITY. *Information, Communication & Society* 2011 Sep;14(6):850-871. [doi: [10.1080/1369118X.2011.587882](#)]
11. McKee M, Cole K, Hurst L, Aldridge R, Horton R. The other Twitter revolution: how social media are helping to monitor the NHS reforms. *BMJ* 2011;342:d948. [Medline: [21325389](#)]
12. Coiera E. Social networks, social media, and social diseases. *BMJ* 2013;346:f3007. [Medline: [23697672](#)]

13. 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](https://doi.org/10.1371/journal.pone.0014118)] [Medline: [21124761](https://pubmed.ncbi.nlm.nih.gov/21124761/)]
14. Scanfeld D, Scanfeld V, Larson EL. Dissemination of health information through social networks: twitter and antibiotics. *Am J Infect Control* 2010 Apr;38(3):182-188 [FREE Full text] [doi: [10.1016/j.ajic.2009.11.004](https://doi.org/10.1016/j.ajic.2009.11.004)] [Medline: [20347636](https://pubmed.ncbi.nlm.nih.gov/20347636/)]
15. Bick D. Media portrayal of birth and the consequences of misinformation. *Midwifery* 2010 Apr;26(2):147-148. [doi: [10.1016/j.midw.2010.02.002](https://doi.org/10.1016/j.midw.2010.02.002)] [Medline: [20206424](https://pubmed.ncbi.nlm.nih.gov/20206424/)]
16. Kelley MC, Trinidad SB. Silent loss and the clinical encounter: Parents' and physicians' experiences of stillbirth-a qualitative analysis. *BMC Pregnancy Childbirth* 2012;12:137 [FREE Full text] [doi: [10.1186/1471-2393-12-137](https://doi.org/10.1186/1471-2393-12-137)] [Medline: [23181615](https://pubmed.ncbi.nlm.nih.gov/23181615/)]
17. Central Statistics Office. Population and Migration Estimates URL: <http://www.cso.ie/en/releasesandpublications/er/pme/populationandmigrationestimatesapril2014/> [accessed 2016-06-17] [WebCite Cache ID [6iKwRU1uH](https://www.webcitation.org/6iKwRU1uH)]
18. Kennedy P. Healthcare reform: maternity service provision in Ireland. *Health Policy* 2010 Oct;97(2-3):145-151. [doi: [10.1016/j.healthpol.2010.04.002](https://doi.org/10.1016/j.healthpol.2010.04.002)] [Medline: [20483499](https://pubmed.ncbi.nlm.nih.gov/20483499/)]
19. Healthcare Pricing Office (HPO), Health Service Executive (HSE). 2014. Perinatal Statistics Report 2013 URL: [https://www.hse.ie/eng/services/publications/hospitals/Perinatal\\_Statistics\\_Report\\_2013.pdf](https://www.hse.ie/eng/services/publications/hospitals/Perinatal_Statistics_Report_2013.pdf) [accessed 2016-06-21] [WebCite Cache ID [6iQxUnEIF](https://www.webcitation.org/6iQxUnEIF)]
20. Manning E, Corcoran P, Meaney S, Greene RA. Perinatal Mortality in Ireland Annual Report 2014. Cork: National Perinatal Epidemiology Centre, 2016 on behalf of the Perinatal Mortality Group URL: <https://www.ucc.ie/en/media/research/nationalperinatalepidemiologycentre/NPECMortality2014130616.pdf> [accessed 2016-06-21] [WebCite Cache ID [6iQxowF7M](https://www.webcitation.org/6iQxowF7M)]
21. RTE About. How RTÉ Is Run URL: <http://www.rte.ie/about/en/how-rte-is-run/2014/03/10/601244-about-rte/> [accessed 2016-06-21] [WebCite Cache ID [6iQy0AKON](https://www.webcitation.org/6iQy0AKON)]
22. Holohan T. Health Service Executive, Ireland. 2014. HSE Midland Regional Hospital, Portlaoise Perinatal Deaths (2006-date) URL: [http://health.gov.ie/wp-content/uploads/2014/03/portlaoise\\_perinatal\\_deaths.pdf](http://health.gov.ie/wp-content/uploads/2014/03/portlaoise_perinatal_deaths.pdf) [accessed 2016-06-21] [WebCite Cache ID [6iQzz92BK](https://www.webcitation.org/6iQzz92BK)]
23. Downs A. Public Interest. 1972. Up and down with ecology - the "issue-attention cycle" URL: [https://www.unc.edu/~fbaum/teaching/articles/Downs\\_Public\\_Interest\\_1972.pdf](https://www.unc.edu/~fbaum/teaching/articles/Downs_Public_Interest_1972.pdf) [accessed 2016-06-21] [WebCite Cache ID [6iQzppsxy](https://www.webcitation.org/6iQzppsxy)]
24. Donelle L, Booth R. Health tweets: an exploration of health promotion on twitter. *Online J Issues Nurs* 2012 Sep;17(3):4 [FREE Full text] [Medline: [23036060](https://pubmed.ncbi.nlm.nih.gov/23036060/)]
25. Procter R, Vis F, Voss A. Reading the riots on Twitter: methodological innovation for the analysis of big data. *International Journal of Social Research Methodology* 2013 May;16(3):197-214. [doi: [10.1080/13645579.2013.774172](https://doi.org/10.1080/13645579.2013.774172)]
26. Braun V, Clarke V. Using thematic analysis in psychology. *Qualitative Research in Psychology* 2006 Jan;3(2):77-101. [doi: [10.1191/1478088706qp063oa](https://doi.org/10.1191/1478088706qp063oa)]
27. Burnap P, Williams M, Sloan L, Rana O, Housley W, Edwards A, et al. Tweeting the terror: modelling the social media reaction to the Woolwich terrorist attack. *Soc. Netw. Anal. Min* 2014 Jun 13;4(1):1-4. [doi: [10.1007/s13278-014-0206-4](https://doi.org/10.1007/s13278-014-0206-4)]
28. Chou W, Hunt Y, Beckjord E, Moser R, Hesse B. Social media use in the United States: implications for health communication. *J med Internet res* 2009;11(4). [doi: [10.2196/jmir.1249](https://doi.org/10.2196/jmir.1249)]
29. Gray NJ, Klein JD, Noyce PR, Sesselberg TS, Cantrill JA. Health information-seeking behaviour in adolescence: the place of the internet. *Soc Sci Med* 2005 Apr;60(7):1467-1478. [doi: [10.1016/j.socscimed.2004.08.010](https://doi.org/10.1016/j.socscimed.2004.08.010)] [Medline: [15652680](https://pubmed.ncbi.nlm.nih.gov/15652680/)]
30. Coleman S, Anthony S, Morrison D. Reuters Institute for the Study of Journalism, University of Oxford. Public trust in the news: a constructivist study of the social life of the news URL: [http://reutersinstitute.politics.ox.ac.uk/sites/default/files/Public%20Trust%20in%20the%20News%20A%20Constructivist%20Study%20of%20the%20Social%20Life%20of%20the%20News\\_0.pdf](http://reutersinstitute.politics.ox.ac.uk/sites/default/files/Public%20Trust%20in%20the%20News%20A%20Constructivist%20Study%20of%20the%20Social%20Life%20of%20the%20News_0.pdf) [accessed 2016-06-21] [WebCite Cache ID [6iQytgu5w](https://www.webcitation.org/6iQytgu5w)]
31. Peterson G, Aslani P, Williams K. How do consumers search for and appraise information on medicines on the Internet? A qualitative study using focus groups. *J med internet res* 2003;5(4). [doi: [10.2196/jmir.5.4.e33](https://doi.org/10.2196/jmir.5.4.e33)]
32. Expectant mothers shun hospital after safety fears. *The Independent*; August 5th 2015 URL: <http://www.independent.ie/irish-news/health/expectant-mothers-shun-hospital-after-safety-fears-31428564.html> [accessed 2016-06-21] [WebCite Cache ID [6iQyUPLcb](https://www.webcitation.org/6iQyUPLcb)]
33. Thackeray R, Neiger BL, Smith AK, Van Wagenen Sarah B. Adoption and use of social media among public health departments. *BMC Public Health* 2012;12:242 [FREE Full text] [doi: [10.1186/1471-2458-12-242](https://doi.org/10.1186/1471-2458-12-242)] [Medline: [22449137](https://pubmed.ncbi.nlm.nih.gov/22449137/)]

## Abbreviations

**RTE:** Raidió Teilifís Éireann



*Edited by G Eysenbach; submitted 10.11.15; peer-reviewed by J Sharpe, K Weller, C Aniftos, S Liang, T Hernández; comments to author 03.04.16; revised version received 24.05.16; accepted 10.06.16; published 27.07.16.*

*Please cite as:*

*Meaney S, Cussen L, Greene RA, O'Donoghue K*

*Reaction on Twitter to a Cluster of Perinatal Deaths: A Mixed Method Study*

*JMIR Public Health Surveill 2016;2(2):e36*

*URL: <http://publichealth.jmir.org/2016/2/e36/>*

*doi: [10.2196/publichealth.5333](https://doi.org/10.2196/publichealth.5333)*

*PMID: [27466002](https://pubmed.ncbi.nlm.nih.gov/27466002/)*

©Sarah Meaney, Leanne Cussen, Richard A Greene, Keelin O'Donoghue. Originally published in JMIR Public Health and Surveillance (<http://publichealth.jmir.org>), 27.07.2016. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Public Health and Surveillance, is properly cited. The complete bibliographic information, a link to the original publication on <http://publichealth.jmir.org>, as well as this copyright and license information must be included.



Original Paper

# Chiropractic and Spinal Manipulation Therapy on Twitter: Case Study Examining the Presence of Critiques and Debates

Alessandro R Marcon<sup>1</sup>, MA (AppLing); Philip Klostermann<sup>2</sup>; Timothy Caulfield<sup>1</sup>, LLM

<sup>1</sup>Health Law Institute, Department of Law, University of Alberta, Edmonton, AB, Canada

<sup>2</sup>The School of Computer Science, Carleton University, Ottawa, ON, Canada

**Corresponding Author:**

Timothy Caulfield, LLM

Health Law Institute

Department of Law

University of Alberta

University Campus, NW

Edmonton, AB, T6G 2H5

Canada

Phone: 1 780 492 9575

Fax: 1 780 492 9575

Email: [caulfield@ualberta.ca](mailto:caulfield@ualberta.ca)

## Abstract

**Background:** Spinal manipulation therapy (SMT) is a popular though controversial practice. The debates surrounding efficacy and risk of SMT are only partially evident in popular discourse.

**Objective:** This study aims to investigate the presence of critiques and debates surrounding efficacy and risk of SMT on the social media platform Twitter. The study examines whether there is presence of debate and whether critical information is being widely disseminated.

**Methods:** An initial corpus of 31,339 tweets was compiled through Twitter's Search Application Programming Interface using the query terms "chiropractic," "chiropractor," and "spinal manipulation therapy." Tweets were collected for the month of December 2015. Post removal of tweets made by bots and spam, the corpus totaled 20,695 tweets, of which a sample (n=1267) was analyzed for skeptical or critical tweets. Additional criteria were also assessed.

**Results:** There were 34 tweets explicitly containing skepticism or critique of SMT, representing 2.68% of the sample (n=1267). As such, there is a presence of 2.68% of tweets in the total corpus, 95% CI 0-6.58% displaying explicitly skeptical or critical perspectives of SMT. In addition, there are numerous tweets highlighting the health benefits of SMT for health issues such as attention deficit hyperactivity disorder (ADHD), immune system, and blood pressure that receive scant critical attention. The presence of tweets in the corpus highlighting the risks of "stroke" and "vertebral artery dissection" is also minute (0.1%).

**Conclusions:** In the abundance of tweets substantiating and promoting chiropractic and SMT as sound health practices and valuable business endeavors, the debates surrounding the efficacy and risks of SMT on Twitter are almost completely absent. Although there are some critical voices of SMT proving to be influential, issues persist regarding how widely this information is being disseminated.

(*JMIR Public Health Surveill* 2016;2(2):e153) doi:[10.2196/publichealth.5739](https://doi.org/10.2196/publichealth.5739)

## KEYWORDS

spinal manipulation; manipulation therapy; chiropractic; alternative medicine; Twitter; social media; infodemiology

## Introduction

Despite its popularity, spinal manipulation therapy (SMT) remains a controversial practice in some circles [1,2]. While there are studies that suggest it is effective for some conditions—such as lower back pain [3-5]—other studies have

questioned its clinical value or have found it to be no more effective than other approaches [6-11]. The issue of risk has also stirred debates. While some believe the concerns are overstated [12-15], other commentators point to possible serious health risks associated with treatment [16-23]. Complications are also evident in terms of establishing and evaluating the

safety culture of SMT [24]. Further adding to the debate is the fact that many chiropractic clinics make claims about the health benefits of SMT for ailments for which there is little to no evidence such as for the treatment of attention deficit hyperactivity disorder (ADHD), asthma, and the boosting of the immune system [25-28]. Although data are uncertain on exactly how many people are treated for the mentioned conditions, these kinds of marketing claims are common and have created policy challenges throughout the world [26,28-30].

In this study, we explore how SMT and related controversies are addressed on social media. A growing body of literature has considered the impact and growing significance of social media, such as Twitter, as a source of health information for the general public [31-34]. Exactly how health information gets disseminated and how people are affected by that information, however, remains complex even though there seems to be little doubt that social media play an increasingly significant role [35,36]. On one hand, it has been noted that social media can function as a democratic, information-disseminating tool which increases the exposure to new information and diverse perspectives [37]. On the other hand, some recent studies have demonstrated how online social networks can come to be structured on the basis of social homophily [38], whereby individuals increasingly and primarily interact with others similar to themselves [39-41]. As a result, studies have shown that using social media can limit the diversity of one's news [42,43] as well as create political polarization over contentious issues, especially on Twitter [44-46].

Although studies show that social media can expose individuals to novel information and diverse opinions [40,43,47], it also has the potential to create "filter bubbles" or "echo chambers," structured with ever-increasing personalization algorithms, in which one's views and perspectives are more often reinforced than called into question [48-50]. Because having limited access to varying perspectives can play a role in augmenting validation and confirming potential bias despite the presence of contradictory information, [51-53] questions are raised as to how positions on a health topic such as SMT are formed, held, reinforced, and contested. Understanding how chiropractic and SMT is portrayed on Twitter, therefore, will provide insight into both the salience of the efficacy or risk debate, the public understanding and awareness of the associated controversies, and the manner in which health information is disseminated.

## Methods

In order to capture public discussions about SMT (eg, "spinal manipulation therapy" is a technical term unlikely to be used by the public on Twitter, particularly given the 140 character limit), we used broad search criteria including the terms "chiropractic," "chiropractor," and "spinal manipulation" on Twitter's Search API. As written on Twitter's frequently asked questions, API stands for "Application Programming Interface" and "is a defined way for a program to accomplish a task, usually by retrieving or modifying data" [54]. In practical terms, Twitter provides the programming code structuring their media, which allows users to design and build software that interacts with Twitter and its data. Our team designed a program that

interacts with Twitter's API search engine allowing us to input search terms, then collect and store all tweets in which those terms appeared. Although very large datasets can be built through Twitter's API, it is not possible to obtain every tweet matching the search criteria. Instead, tweets deemed most "relevant" are provided by Twitter.

An initial search revealed over 3,000,000 tweets, so we selected the most recent full month: December 2015, and created 3 corpora based on each of the terms mentioned. "Spinal manipulation" is a term used frequently by health care providers but not the general public, so for the purpose of this study, the more inclusive terms "chiropractic" and "chiropractor" were also searched to access more general Twitter discourse. Important to note is that searching for tweets with the terms "chiropractic" and "chiropractor" also captured these terms prefixed with a hashtag (#), a key component of Twitter communication [32,47,55]. Data collection of each tweet included: username, twitter handle, tweet, number of retweets, number of likes, time and date, and city location (if the Twitter user had included geo-tags). The data collection resulted in the following number of tweets (including retweets): Corpus 1, "chiropractic," 18,354 tweets; Corpus 2, "chiropractor," 12,918 tweets; and Corpus 3, "spinal manipulation," 67 tweets.

An initial exploratory analysis of approximately 600 tweets in each of Corpora 1 and 2, as well as the entire Corpus 3 was conducted in order to identify general themes in the discourse [56]. Next, spam was deleted from Corpora 1 and 2 by highlighting very active users (more than 15 tweets in the month) and removing accounts deemed to be bots, which are automated (robotic) accounts programmed to perform simple, repetitive tasks on social media. On Twitter, a bot might tweet to provide links to a wide range of promotional material, follow accounts that other accounts follow, retweet others' tweets based on key words, or tweet nonsensical phrases with embedded key words. Bots typically have no bios, tweet extensively and periodically (eg, every hour), and are seldom followed by human users. In addition, because there were a large number of bots spreading promotional spam in the corpus, all usernames with the words "job" or "deal" were removed on the grounds of being promotional spam bots. Although spam does provide data worthy of analysis, real and active users with a large body of "followers" are much more influential [57]. Following the removal of most discernible spam and bots, the final number of tweets in each corpus is as followed: Corpus 1, "chiropractic," 11,446 tweets; Corpus 2, "chiropractor," 9182 tweets; and Corpus 3, "spinal manipulation," 67 tweets.

Using a confidence level of 95% and a CI of 3.9, it was determined that a sample of 600 tweets, sequentially organized by date were to be analyzed in Corpora 1 and 2 (n=1200). All tweets were analyzed in Corpus 3 (n=67). Tweets were deemed skeptical or critical if they raised any doubts of efficacy, highlighted potential health risks, mentioned excessive pain, labeled the treatment with negative, derogatory terms, linked SMT to criminal activity or questionable health practices, or asked questions concerning efficacy that were, at times, followed by links. After all skeptical or critical tweets were identified, each tweet was analyzed with greater scrutiny (opening links, viewing emojis and pictures, and assessing the context of the

dialogue) and was identified as explicitly containing skeptical or critical views of SMT. Because textual analysis is often understood to be a subjective process, 50% of the total tweets in the sample (n=600) were tested for inter-coder reliability using Cohen kappa, resulting in  $\kappa=.95$ . This Kappa score indicates almost perfect inter-rater agreement according to Landis and Koch's benchmark standards [58].

To shed light on how the tweets were disseminated, the number of mentions and hashtags in each critical or skeptical tweet were counted. A mention is a Twitter tool whereby a tweet contains the "@" sign, followed by a username of another Twitter account. This user who is mentioned is notified of appearing in another person's tweet. Using mentions often creates dialogue between twitter users. A hashtag, "#," followed by a key word or phrase is a way to create an information category in which relevant information can be appended. Hashtags can be searched on Twitter to view all tweets constituting a particular category.

Next, the presence of skeptical or critical tweets were assessed by assembling lists of the top 10 retweets and liked tweets in each corpus. Assessing retweets and likes highlights how much attention particular tweets have been given and illustrates which information is most widely disseminated using Twitter tools. In addition, in order to explore how controversial applications of SMT are represented, all tweets containing "ADHD," "immune system," and "blood pressure" (all applications with evidence to support the use of SMT) were highlighted and examined. The objective was to determine if there were tweets critical of claims suggesting chiropractic or SMT can benefit ailments pertaining to these health issues. Finally, all tweets with the key words: "vertebral artery dissection" or "stroke," were identified and analyzed, as these terms, highlighted in the

relevant literature on risk, are indicative of potential risks associated with SMT [13,21,22].

## Results

Of all tweets analyzed in Corpora 1 and 2 (n=1200), a total of 77 tweets (6.42%), 95% CI (2.52%-10.32%) contained skeptical or critical sentiment. Following in-depth analysis, 25 of the 77 tweets contained explicitly skeptical or critical content, representing 2.08% of the more general Twitter discourse, 95% CI (0%-5.98%). In Corpus 3: "spinal manipulation" (n=67), 25 tweets, 37% of the corpus, contained skeptical or critical sentiments. Following in-depth analysis, 9 of the 25 tweets contained explicitly skeptical or critical content, representing 13% of the Corpus. Of the 34 total skeptical or critical tweets, a total of 7 contained mentions (21%) and 5 contained hashtags (15%). For examples of skeptical or critical tweets, refer to [Figure 1](#). To view all skeptical or critical tweets, refer to [Multimedia Appendix 1](#).

Regarding benefits for specific health issues, a total of 88 tweets of 20,695 mentioned the terms "ADHD," "immune system," or "blood pressure." Of those 88 tweets, 4 (5%) were identified as skeptical or critical ([Table 1](#)). These 4 tweets can be found in the [Multimedia Appendix 1](#). In terms of highlighting specific risks associated with SMT, of 20,695 tweets, 30 (0.14%) mentioned "stroke" or "vertebral artery dissection." Of these 30 tweets, 22 (73%) contained explicitly skeptical or critical content ([Table 2](#)). Of these 22 tweets, all unique skeptical or critical tweets can be found in the [Multimedia Appendix 1](#). Finally, regarding tweet impact in their respective Corpus, 4 skeptical or critical tweets ranked in the top 10 for retweets, whereas 1 skeptical or critical tweet ranked in the top 10 for likes ([Figure 2](#)). All these tweets were found in Corpus 1.

**Table 1.** Tweets and skeptical or critical tweets for health issues.

Health issue term	Tweets containing term	Skeptical or critical tweets containing term
Attention deficit hyperactivity disorder	18	3
Immune system	44	1
Blood pressure	26	0
Total	88	4

**Table 2.** Tweets and skeptical or critical tweets highlighting risks of spinal manipulation therapy.

Term associated with risk	Total tweets containing term	Skeptical or critical tweet containing term
Stroke	10	4
Vertebral artery dissection	20	18
Total	30	22

**Figure 1.** Examples of skeptical or critical tweets with usernames replaced by @#####.

Term in tweet	Example tweets
Chiropractic	<p>Thinking about chiropractic thyroid program? Here's @##### 's take  <a href="http://abt.cm/1ZUYvlu">http://abt.cm/1ZUYvlu</a> <a href="http://pic.twitter.com/HNYAruWVao">pic.twitter.com/HNYAruWVao</a></p> <p>What did not cure my infection: Magic, organic food, chiropractic, homeopathy, acupuncture.\n\nWhat did cure my infection: Big pharma.</p> <p>@##### @##### ASRF holds charity status. This must end #StopAVN #auspol #chiropractic @#####  <a href="https://twitter.com/SueIeraci/status/681287717049597952">https://twitter.com/SueIeraci/status/681287717049597952</a> ...Vitalistic            Chiro is not science-based health care  <a href="https://spinalresearch.com.au/wp-content/uploads/ILLUMINATE-415-FINAL.pdf">https://spinalresearch.com.au/wp-content/uploads/ILLUMINATE-415-FINAL.pdf</a> ... <a href="http://pic.twitter.com/7JhVEoxnOm">pic.twitter.com/7JhVEoxnOm</a></p>
Chiropractor	<p>Just once, I'd love to see a chiropractor who got busted for doing something unprofessional say "Fair cop, gov. You got me. I was wrong"</p> <p>Going to the chiropractor actually made my back worse... Is that possible?</p> <p>Everytime I go to the chiropractor I end up in more pain than before ):</p>
Spinal manipulation	<p>Improper Chiropractic spinal manipulation has been linked with increased risk of stroke. We can help. <a href="http://bit.ly/1AD9nLr">http://bit.ly/1AD9nLr</a></p> <p>Spinal manipulation/mobilisation for neck pain: caution is warranted  <a href="http://edzardernst.com/2015/12/spinal-manipulationmobilisation-for-neck-pain-caution-is-warranted/">http://edzardernst.com/2015/12/spinal-manipulationmobilisation-for-neck-pain-caution-is-warranted/</a> ... via @#####</p> <p>No change in cervical lordosis after spinal manipulation [prospective trial]. Surprise, surprise!  <a href="http://chiromt.biomedcentral.com/articles/10.1186/s12998-015-0078-3">http://chiromt.biomedcentral.com/articles/10.1186/s12998-015-0078-3</a></p>



**Figure 2.** Skeptical or critical tweets ranking in the top 10 for retweets and likes.

Tweet	Number of retweets	Retweet rank in corpus	Number of likes	Like rank in corpus	Corpus
Vertebral artery dissection after chiropractic manipulation: yet another case <a href="http://edzardernst.com/2015/12/vertebral-artery-dissection-after-chiropractic-manipulation-yet-another-case/">http://edzardernst.com/2015/12/vertebral-artery-dissection-after-chiropractic-manipulation-yet-another-case/</a> ... via @#####	21	8 <sup>th</sup> (tie)	18	Not in Top 10	1
Low vax rates in Manitoba lead to #pertussis outbreaks   At center of #chiropractic anti-vax lobby. #VaccinesWork <a href="http://www.winnipegfreepress.com/local/whopping-cough-alarm-sounded-363062481.html">http://www.winnipegfreepress.com/local/whopping-cough-alarm-sounded-363062481.html</a>	21	8 <sup>th</sup> (tie)	8	Not in Top 10	1
Things to remove from the Medicare Benefits Schedule: Chiropractic.	17	9 <sup>th</sup> (tie)	25	10 <sup>th</sup> (tie)	1
"Manitoba is regarded in some areas as the centre of the chiropractic anti-immunization lobby." <a href="http://www.winnipegfreepress.com/local/whopping-cough-alarm-sounded-363062481.html">http://www.winnipegfreepress.com/local/whopping-cough-alarm-sounded-363062481.html</a> ...	17	9 <sup>th</sup> (tie)	6	Not in Top-10	1

## Discussion

### Principal Findings

This analysis of social media finds that the efficacy of chiropractic and SMT is rarely doubted or questioned on Twitter. In addition, the potential risks are rarely mentioned or debated. The manner in which efficacy and risks are tweeted across the 3 corpora, based on the different search terms, however, reveals some insights regarding how contentious or contrary information is, or can be, disseminated on the social media platform.

“Spinal manipulation” is a more specialized term and as a result seems likely to be associated with more technical twitter discussions. As evident in [Multimedia Appendix 1](#), of the 9 skeptical or critical tweets using “spinal manipulation” include a link to academic studies or research. Tweeters using this term, therefore, seem more aware of the debates surrounding the efficacy and potential risks of SMT.

Although both terms “chiropractor” and “chiropractic” yielded very few skeptical or critical tweets, some of the skeptical or critical tweets using “chiropractic” had been liked and retweeted significantly. This suggests that some skeptical or critical

perspectives have an impact on the tweeting public despite the fact that their voices are marginal in number. A question arises, however, as to whether this information is simply being liked and retweeted by like-minded individuals inside of a social network bubble or if this information is reaching new audiences [48]. Even the discussions around highly controversial uses of SMT—as highlighted by our analysis of the Tweets associated with asthma, ADHD, and the immune system—did not generate significant critical attention.

Studies have shown that group polarization is prevalent in Twitter conversations involving politics and contentious issues [44,45,59], thereby limiting information dissemination among those with opposing views. It is yet to be explored, however, as to whether something similar occurs in health-related discourse on Twitter. If it is the case that the increasingly personalized algorithms structuring the Internet expose individuals more often to information that reinforces one’s view and less often to novel information [48], it’s possible that heuristics like the confirmation bias are being magnified [51-53]. The degree to which information is shared among dissimilar individuals on social media regarding less politicized topics, such as health, still requires further research [59]. Regardless,



on Twitter, studies have shown that hashtags and mentions can prove to be useful tools for disseminating information more widely and for engaging more diverse audiences [45,47]. In our study, of the 34 tweets explicitly expressing skepticism or doubt in the sampled tweets, only 7 included mentions and 5 included hashtags. This arguably demonstrates a narrow scope of information dissemination. In short, those in the health community wishing to make their critiques of chiropractic and SMT better known to a broader public might find using mentions and hashtags beneficial to their cause—especially because of the impact that Twitter can have on the formation of views is well-documented [35,36].

### Limitations

This study has several limitations worth noting. Given the nature of Twitter discussions and the somewhat limited access provided by Twitter's API, it can be challenging to capture a comprehensive collection of tweets on any topic. In addition, other potential terms such as “chiro” and “spinal adjustment”

are present on Twitter, which may produce datasets with somewhat different results. Finally, although December 2015 was chosen at random, there is nothing to suggest that other time frames would be significantly similar or different. Despite these limitations, this study highlights the degree to which discussions of risk and critical views on efficacy are almost completely absent from Twitter.

### Conclusion

In total, Twitter representations of SMT and chiropractic are overwhelmingly not skeptical or critical. The ongoing debates regarding efficacy and risk in the academic literature and the popular press [60-62] do not have a strong presence on Twitter. This study provides insight into how Twitter users discuss SMT and chiropractic and suggests that, in the aggregate, the information on this social media platform is far from balanced or informed. Although voices do exist which raise concerns of SMT efficacy and highlight potential risks associated with the practice, their presence is marginal in overall Twitter discourse.

### Acknowledgments

In addition to funding support from the Trudeau Foundation, this work was supported by funding from the Canadian Institutes of Health Research, Alberta Innovates-Health Solutions, and the Women and Children's Health Research Institute, University of Alberta. The authors also thank Christen Rachul for providing assistance and guidance. They appreciate Robyn Hyde-Lay for her various contributions during the manuscript preparation and submission process, and the University of Alberta's Health Law Institute for its administrative support.

### Authors' Contributions

ARM and TC designed the study. PK built the program to collect the dataset. ARM developed the methodology with input from TC. ARM performed the data analysis. ARM drafted the manuscript with input from TC. ARM and TC collaborated in editing and revising the manuscript.

### Conflicts of Interest

None declared.

### Multimedia Appendix 1

All skeptical and critical tweets.

[[PDF File \(Adobe PDF File\), 34KB - publichealth\\_v2i2e153\\_app1.pdf](#)]

### References

1. Homola S. Science-Based Medicine. 2015. Opposing chiropractic: persecution or justified criticism URL: <https://www.sciencebasedmedicine.org/opposing-chiropractic-persecution-or-justified-criticism/> [accessed 2016-09-09] [WebCite Cache ID 6kOxIkZlB]
2. Ingraham P. PainScience.com. 2016 Jul 02. Does chiropractic work? URL: <https://www.painscience.com/articles/does-chiropractic-work.php> [accessed 2016-09-09] [WebCite Cache ID 6kOxLDdta]
3. Santilli V, Beghi E, Finucci S. Chiropractic manipulation in the treatment of acute back pain and sciatica with disc protrusion: a randomized double-blind clinical trial of active and simulated spinal manipulations. *Spine J* 2006;6(2):131-137. [doi: [10.1016/j.spinee.2005.08.001](https://doi.org/10.1016/j.spinee.2005.08.001)] [Medline: [16517383](#)]
4. Goertz CM, Long CR, Hondras MA, Petri R, Delgado R, Lawrence DJ, et al. Adding chiropractic manipulative therapy to standard medical care for patients with acute low back pain: results of a pragmatic randomized comparative effectiveness study. *Spine (Phila Pa 1976)* 2013 Apr 15;38(8):627-634. [doi: [10.1097/BRS.0b013e31827733e7](https://doi.org/10.1097/BRS.0b013e31827733e7)] [Medline: [23060056](#)]
5. Meade TW, Dyer S, Browne W, Townsend J, Frank AO. Low back pain of mechanical origin: randomised comparison of chiropractic and hospital outpatient treatment. *BMJ* 1990 Jun 2;300(6737):1431-1437 [FREE Full text] [Medline: [2143092](#)]

6. Rubinstein SM, Terwee CB, Assendelft WJ, de Boer MR, van Tulder MW. Spinal manipulative therapy for acute low back pain: an update of the cochrane review. *Spine (Phila Pa 1976)* 2013 Feb 1;38(3):E158-E177. [doi: [10.1097/BRS.0b013e31827dd89d](https://doi.org/10.1097/BRS.0b013e31827dd89d)] [Medline: [23169072](https://pubmed.ncbi.nlm.nih.gov/23169072/)]
7. Dougherty PE, Karuza J, Dunn AS, Savino D, Katz P. Spinal Manipulative Therapy for Chronic Lower Back Pain in Older Veterans: A Prospective, Randomized, Placebo-Controlled Trial. *Geriatr Orthop Surg Rehabil* 2014 Dec;5(4):154-164 [FREE Full text] [doi: [10.1177/2151458514544956](https://doi.org/10.1177/2151458514544956)] [Medline: [26246937](https://pubmed.ncbi.nlm.nih.gov/26246937/)]
8. Crothers AL, French SD, Hebert JJ, Walker BF. Spinal manipulative therapy, Graston technique® and placebo for non-specific thoracic spine pain: a randomised controlled trial. *Chiropractic & Manual Therapies* 2016;24:16. [doi: [10.1186/s12998-016-0096-9](https://doi.org/10.1186/s12998-016-0096-9)]
9. Posadzki P, Ernst E. Spinal manipulation: an update of a systematic review of systematic reviews. *N Z Med J* 2011 Aug 12;124(1340):55-71. [Medline: [21952385](https://pubmed.ncbi.nlm.nih.gov/21952385/)]
10. Wong AY, Parent EC, Dhillon SS, Prasad N, Kawchuk GN. Do participants with low back pain who respond to spinal manipulative therapy differ biomechanically from nonresponders, untreated controls or asymptomatic controls? *Spine (Phila Pa 1976)* 2015 Sep 1;40(17):1329-1337. [doi: [10.1097/BRS.0000000000000981](https://doi.org/10.1097/BRS.0000000000000981)] [Medline: [26020851](https://pubmed.ncbi.nlm.nih.gov/26020851/)]
11. Merepeza A. Effects of spinal manipulation versus therapeutic exercise on adults with chronic low back pain: a literature review. *J Can Chiropr Assoc* 2014 Dec;58(4):456-466. [Medline: [25550671](https://pubmed.ncbi.nlm.nih.gov/25550671/)]
12. Cassidy JD, Bronfort G, Hartvigsen J. Should we abandon cervical spine manipulation for mechanical neck pain? *No. BMJ* 2012;344:e3680. [Medline: [22677797](https://pubmed.ncbi.nlm.nih.gov/22677797/)]
13. Cassidy JD, Boyle E, Côté P, He Y, Hogg-Johnson S, Silver FL, et al. Risk of vertebral stroke and chiropractic care: results of a population-based case-control and case-crossover study. *J Manipulative Physiol Ther* 2009 Feb;32(2 Suppl):S201-S208. [doi: [10.1016/j.jmpt.2008.11.020](https://doi.org/10.1016/j.jmpt.2008.11.020)] [Medline: [19251066](https://pubmed.ncbi.nlm.nih.gov/19251066/)]
14. Kosloff TM, Elton D, Tao J, Bannister WM. Chiropractic care and the risk of vertebral stroke: results of a case-control study in U.S. commercial and Medicare Advantage populations. *Chiropr Man Therap* 2015;23:19 [FREE Full text] [doi: [10.1186/s12998-015-0063-x](https://doi.org/10.1186/s12998-015-0063-x)] [Medline: [26085925](https://pubmed.ncbi.nlm.nih.gov/26085925/)]
15. Thiel HW, Bolton JE, Docherty S, Portlock JC. Safety of chiropractic manipulation of the cervical spine: a prospective national survey. *Spine (Phila Pa 1976)* 2007 Oct 1;32(21):2375-8; discussion 2379. [doi: [10.1097/BRS.0b013e3181557bb1](https://doi.org/10.1097/BRS.0b013e3181557bb1)] [Medline: [17906581](https://pubmed.ncbi.nlm.nih.gov/17906581/)]
16. Mann T, Refshauge KM. Causes of complications from cervical spine manipulation. *Aust J Physiother* 2001;47(4):255-266 [FREE Full text] [Medline: [11722294](https://pubmed.ncbi.nlm.nih.gov/11722294/)]
17. Rivett DA, Milburn P. Complications Arising from Spinal Manipulative Therapy in New Zealand. *Physiotherapy* 1997 Dec;83(12):626-632. [doi: [10.1016/S0031-9406\(05\)65929-9](https://doi.org/10.1016/S0031-9406(05)65929-9)]
18. Haynes MJ, Vincent K, Fischhoff C, Bremner AP, Lanlo O, Hankey GJ. Assessing the risk of stroke from neck manipulation: a systematic review. *Int J Clin Pract* 2012 Oct;66(10):940-947 [FREE Full text] [doi: [10.1111/j.1742-1241.2012.03004.x](https://doi.org/10.1111/j.1742-1241.2012.03004.x)] [Medline: [22994328](https://pubmed.ncbi.nlm.nih.gov/22994328/)]
19. Wand BM, Heine PJ, O'Connell NE. Should we abandon cervical spine manipulation for mechanical neck pain? Yes. *BMJ* 2012;344:e3679. [Medline: [22677796](https://pubmed.ncbi.nlm.nih.gov/22677796/)]
20. Whedon JM, Song Y, Mackenzie TA, Phillips RB, Lukovits TG, Lurie JD. Risk of stroke after chiropractic spinal manipulation in medicare B beneficiaries aged 66 to 99 years with neck pain. *J Manipulative Physiol Ther* 2015 Feb;38(2):93-101 [FREE Full text] [doi: [10.1016/j.jmpt.2014.12.001](https://doi.org/10.1016/j.jmpt.2014.12.001)] [Medline: [25596875](https://pubmed.ncbi.nlm.nih.gov/25596875/)]
21. Stevinson C, Ernst E. Risks associated with spinal manipulation. *Am J Med* 2002 May;112(7):566-571. [Medline: [12015249](https://pubmed.ncbi.nlm.nih.gov/12015249/)]
22. Reuter U, Hämling M, Kavuk I, Einhäupl KM, Schielke E. Vertebral artery dissections after chiropractic neck manipulation in Germany over three years. *J Neurol* 2006 Jun;253(6):724-730. [doi: [10.1007/s00415-006-0099-x](https://doi.org/10.1007/s00415-006-0099-x)] [Medline: [16511634](https://pubmed.ncbi.nlm.nih.gov/16511634/)]
23. Biller J, Sacco RL, Albuquerque FC, Demaerschalk BM, Fayad P, Long PH, American Heart Association Stroke Council. Cervical arterial dissections and association with cervical manipulative therapy: a statement for healthcare professionals from the american heart association/american stroke association. *Stroke* 2014 Oct;45(10):3155-3174 [FREE Full text] [doi: [10.1161/STR.0000000000000016](https://doi.org/10.1161/STR.0000000000000016)] [Medline: [25104849](https://pubmed.ncbi.nlm.nih.gov/25104849/)]
24. Rozmovits L, Mior S, Boon H. Exploring approaches to patient safety: the case of spinal manipulation therapy. *BMC Complement Altern Med* 2016;16(1):164 [FREE Full text] [doi: [10.1186/s12906-016-1149-2](https://doi.org/10.1186/s12906-016-1149-2)] [Medline: [27251398](https://pubmed.ncbi.nlm.nih.gov/27251398/)]
25. Grod JP, Sikorski D, Keating JC. Unsubstantiated claims in patient brochures from the largest state, provincial, and national chiropractic associations and research agencies. *J Manipulative Physiol Ther* 2001 Oct;24(8):514-519. [doi: [10.1067/mmt.2001.118205](https://doi.org/10.1067/mmt.2001.118205)] [Medline: [11677551](https://pubmed.ncbi.nlm.nih.gov/11677551/)]
26. Sikorski D, Grod J. The unsubstantiated web site claims of chiropractic colleges in Canada and the United States. *J Chiropr Educ* (2) 2003;17:113-119 [FREE Full text]
27. Ferrance RJ, Miller J. Chiropractic diagnosis and management of non-musculoskeletal conditions in children and adolescents. *Chiropr Osteopat* 2010;18:14 [FREE Full text] [doi: [10.1186/1746-1340-18-14](https://doi.org/10.1186/1746-1340-18-14)] [Medline: [20525197](https://pubmed.ncbi.nlm.nih.gov/20525197/)]
28. Shelley J, Clark M, Caulfield T. The face of chiropractic: evidence-based? Focus on Alternative and Complementary Therapies 2014 Nov 18;20(1):13-22. [doi: [10.1111/act.12151](https://doi.org/10.1111/act.12151)]
29. Hanna M, Honeychurch M. Chronic misleading online advertising by chiropractors. *N Z Med J* 2016;129(1432):91-93. [Medline: [27356260](https://pubmed.ncbi.nlm.nih.gov/27356260/)]

30. Harvey K, Vickers M. Doctorportal. 2016 Jan 18. Ken Harvey and Malcolm Vickers: Chiropractic board in firing line URL: <http://www.doctorportal.com.au/mjainsight/2016/1/ken-harvey-and-malcolm-vickers-chiropractic-board-firing-line/> [accessed 2016-02-14] [WebCite Cache ID 6kOxaGbmo]
31. Du L, Kamenova K, Caulfield T. The gene patent controversy on Twitter: a case study of Twitter users' responses to the CHEO lawsuit against Long QT gene patents. *BMC Med Ethics* 2015;16:55 [FREE Full text] [doi: [10.1186/s12910-015-0049-1](https://doi.org/10.1186/s12910-015-0049-1)] [Medline: [26302779](https://pubmed.ncbi.nlm.nih.gov/26302779/)]
32. Radzikowski J, Stefanidis A, Jacobsen KH, Croitoru A, Crooks A, Delamater PL. The Measles Vaccination Narrative in Twitter: A Quantitative Analysis. *JMIR Public Health Surveill* 2016 Jan 04;2(1):e1. [doi: [10.2196/publichealth.5059](https://doi.org/10.2196/publichealth.5059)] [Medline: [27227144](https://pubmed.ncbi.nlm.nih.gov/27227144/)]
33. Cole-Lewis H, Pugatch J, Sanders A, Varghese A, Posada S, Yun C, et al. Social Listening: A Content Analysis of E-Cigarette Discussions on Twitter. *J Med Internet Res* 2015;17(10):e243 [FREE Full text] [doi: [10.2196/jmir.4969](https://doi.org/10.2196/jmir.4969)] [Medline: [26508089](https://pubmed.ncbi.nlm.nih.gov/26508089/)]
34. Paul MJ, Dredze M. JHU. 2011. You are what you tweet: Analyzing twitter for public health URL: [https://www.cs.jhu.edu/~mdredze/publications/twitter\\_health\\_icwsm\\_11.pdf](https://www.cs.jhu.edu/~mdredze/publications/twitter_health_icwsm_11.pdf) [accessed 2016-09-13] [WebCite Cache ID 6kUa3EFqd]
35. Centola D. Social media and the science of health behavior. *Circulation* 2013 May 28;127(21):2135-2144 [FREE Full text] [doi: [10.1161/CIRCULATIONAHA.112.101816](https://doi.org/10.1161/CIRCULATIONAHA.112.101816)] [Medline: [23716382](https://pubmed.ncbi.nlm.nih.gov/23716382/)]
36. 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](https://doi.org/10.1177/1524839911405850)] [Medline: [21558472](https://pubmed.ncbi.nlm.nih.gov/21558472/)]
37. Benkler Y. *The Wealth of Nations: How Social Production Transforms Markets and Freedom*. New Haven, CT: Yale University Press; 2006.
38. McPherson M, Smith-Lovin L, Cook JM. Birds of a Feather: Homophily in Social Networks. *Annu. Rev. Sociol* 2001 Aug;27(1):415-444. [doi: [10.1146/annurev.soc.27.1.415](https://doi.org/10.1146/annurev.soc.27.1.415)]
39. Grevet C, Terveen L, Gilbert E. Managing political differences in social media. 2014 Presented at: Proceedings of the 17th ACM conference on computer supported cooperative work & social computing; Feb ; Baltimore, USA; February 15-19, 2014; Baltimore, MD URL: <http://comp.social.gatech.edu/papers/grevet.cscw14.political.pdf>
40. Bahksy E, Rosenn I, Marlow C, Adamic L. The role of social networks in information diffusion. 2012 Presented at: Proceedings of ACM WWW Session; April 4, 16-20, 2012; Lyon, France p. 16-20 URL: <http://arxiv.org/pdf/1201.4145v2.pdf>
41. Flaxman S, Goel S, Rao JM. Ideological Segregation and the Effects of Social Media on News Consumption. *Public Opinion Quarterly* 2016;80:298-320 [FREE Full text] [doi: [10.2139/ssrn.2363701](https://doi.org/10.2139/ssrn.2363701)]
42. Nikolov D, Oliveira D, Flammini A, Menczer F. Measuring online social bubbles. *Peer J Computer Science* 2015;1:e28. [doi: [10.7717/peerj-cs.38](https://doi.org/10.7717/peerj-cs.38)]
43. Bakshy E, Messing S, Adamic LA. Political science. Exposure to ideologically diverse news and opinion on Facebook. *Science* 2015 Jun 5;348(6239):1130-1132. [doi: [10.1126/science.aaa1160](https://doi.org/10.1126/science.aaa1160)] [Medline: [25953820](https://pubmed.ncbi.nlm.nih.gov/25953820/)]
44. Yardi S, Boyd D. Dynamic Debates: An Analysis of Group Polarization Over Time on Twitter. *Bulletin of Science, Technology & Society* 2010 Sep 14;30(5):316-327. [doi: [10.1177/0270467610380011](https://doi.org/10.1177/0270467610380011)]
45. Conover M, Ratkiewicz J, Francisco M, Goncalves B, Flammini A, Menczer F. Political polarization on Twitter. 2011 Presented at: Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media;; July 17-21, 2011; Barcelona, Spain URL: <https://www.aaai.org/ocs/index.php/ICWSM/ICWSM11/paper/viewFile/2847/3275>
46. Conover MD, Gonçalves B, Flammini A, Menczer F. Partisan asymmetries in online political activity. *EPJ Data Sci* 2012 Jun 18;1:6. [doi: [10.1140/epjds6](https://doi.org/10.1140/epjds6)]
47. Romero D, Meeder B, Kleinberg J. Difference in the mechanics of information diffusion across topics: Idioms, political hashtags, and complex contagion on Twitter. 2011 Presented at: Proceedings of the 20th international conference on WWW; March 28 - April 1, 2011; Hyderabad, India. [doi: [10.1145/1963405.1963503](https://doi.org/10.1145/1963405.1963503)]
48. Pariser E. *The Filter Bubble: How the New Personalized Web Is Changing What We Read and How We Think*. New York, NY: Penguin Books; 2011.
49. Gilbert E, Bergstrom T, Karahalios K. Blogs are echo chambers: blogs are echo chambers. 2009 Presented at: Proceedings of the 42nd Hawaii International Conference on System Sciences , Jan ; Big Island, HI; January 5-8, 2009; Big Island, HI. [doi: [10.1109/HICSS.2009.91](https://doi.org/10.1109/HICSS.2009.91)]
50. Sunstein C. *Republic.com 2.0*. Princeton, NJ: Princeton University Press; 2009.
51. Hart W, Albarracín D, Eagly AH, Brechan I, Lindberg MJ, Merrill L. Feeling validated versus being correct: a meta-analysis of selective exposure to information. *Psychol Bull* 2009 Jul;135(4):555-588. [doi: [10.1037/a0015701](https://doi.org/10.1037/a0015701)] [Medline: [19586162](https://pubmed.ncbi.nlm.nih.gov/19586162/)]
52. Browne M, Thomson P, Rockloff MJ, Pennycook G. Going against the Herd: Psychological and Cultural Factors Underlying the 'Vaccination Confidence Gap'. *PLoS One* 2015;10(9):e0132562 [FREE Full text] [doi: [10.1371/journal.pone.0132562](https://doi.org/10.1371/journal.pone.0132562)] [Medline: [26325522](https://pubmed.ncbi.nlm.nih.gov/26325522/)]
53. Stanovich KE, West RF, Toplak ME. Myside Bias, Rational Thinking, and Intelligence. *Current Directions in Psychological Science* 2013 Aug 05;22(4):259-264. [doi: [10.1177/0963721413480174](https://doi.org/10.1177/0963721413480174)]
54. Twitter. Frequently Asked Questions URL: <https://dev.twitter.com/faq/basics> [accessed 2016-08-04] [WebCite Cache ID 6jWCOkmGy]

55. Chang H. A new perspective on Twitter hashtag use: Diffusion of innovation theory. Proc. Am. Soc. Info. Sci. Tech 2011 Feb 03;47(1):1-4. [doi: [10.1002/meet.14504701295](https://doi.org/10.1002/meet.14504701295)]
56. Saldana, J. The coding manual for qualitative researchers (2nd ed). Los Angeles, CA: SAGE Publications Ltd; 2013.
57. Bahksy E, Hofman J, Mason W, Watts D. Everyone's an influencer: Quantifying influence on Twitter. 2011 Presented at: Proceedings of the 4th ACM International Conference on Web Search and Data Mining; February 9-12, 2011; Kowloon, Hong Kong URL: [http://research.microsoft.com/pubs/164539/b\\_h\\_m\\_w\\_WSDM\\_11.pdf](http://research.microsoft.com/pubs/164539/b_h_m_w_WSDM_11.pdf)
58. Landis J, Koch G. The measurement of observer agreement for categorical data. Biometrics 1977;33:159-174 [FREE Full text] [Medline: [843571](https://pubmed.ncbi.nlm.nih.gov/843571/)]
59. Barberá P, Jost JT, Nagler J, Tucker JA, Bonneau R. Tweeting From Left to Right: Is Online Political Communication More Than an Echo Chamber? Psychol Sci 2015 Oct;26(10):1531-1542. [doi: [10.1177/0956797615594620](https://doi.org/10.1177/0956797615594620)] [Medline: [26297377](https://pubmed.ncbi.nlm.nih.gov/26297377/)]
60. Ernst E. The Guardian. 2012 Apr 27. Chiropractic manipulation of the spine may cause strokes or even death URL: <https://www.theguardian.com/science/blog/2012/apr/27/chiropractic-manipulation-spine-strokes> [accessed 2016-09-09] [WebCite Cache ID 6kOz6fQip]
61. Mohammadi D. The Guardian. 2015 Oct 18. Chiropractic and osteopathy- how do they work? URL: <http://www.theguardian.com/lifeandstyle/2015/oct/18/osteopaths-chiropractors-back-pain-whose-spine-is-it-anyway> [accessed 2016-03-02] [WebCite Cache ID 6fjr9xIVB]
62. Senelick RC. Huffpost Healthy Living. 2012 Oct 24. Chiropractic manipulation can cause a stroke URL: [http://www.huffingtonpost.com/richard-c-senelick-md/chiropractic-treatment-safety\\_b\\_1975979.html](http://www.huffingtonpost.com/richard-c-senelick-md/chiropractic-treatment-safety_b_1975979.html) [accessed 2016-03-01] [WebCite Cache ID 6fjrJYTpm]

## Abbreviations

**ADHD:** attention deficit hyperactivity disorder

**API:** application programming interface

**SMT:** spinal manipulation therapy

*Edited by G Eysenbach; submitted 09.03.16; peer-reviewed by J Hartvigsen, JC Chen; comments to author 05.07.16; revised version received 04.08.16; accepted 21.08.16; published 16.09.16.*

*Please cite as:*

Marcon AR, Klostermann P, Caulfield T

Chiropractic and Spinal Manipulation Therapy on Twitter: Case Study Examining the Presence of Critiques and Debates

JMIR Public Health Surveill 2016;2(2):e153

URL: <http://publichealth.jmir.org/2016/2/e153/>

doi: [10.2196/publichealth.5739](https://doi.org/10.2196/publichealth.5739)

PMID: [27637456](https://pubmed.ncbi.nlm.nih.gov/27637456/)

©Alessandro R Marcon, Philip Klostermann, Timothy Caulfield. Originally published in JMIR Public Health and Surveillance (<http://publichealth.jmir.org>), 16.09.2016. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Public Health and Surveillance, is properly cited. The complete bibliographic information, a link to the original publication on <http://publichealth.jmir.org>, as well as this copyright and license information must be included.



Original Paper

# Building a National Neighborhood Dataset From Geotagged Twitter Data for Indicators of Happiness, Diet, and Physical Activity

Quynh C Nguyen<sup>1</sup>, PhD; Dapeng Li<sup>2</sup>, MS, PhD; Hsien-Wen Meng<sup>1</sup>, MS; Suraj Kath<sup>3</sup>, MS; Elaine Nsoesie<sup>4</sup>, MS, PhD; Feifei Li<sup>3</sup>, PhD; Ming Wen<sup>5</sup>, MA, MS, PhD

<sup>1</sup>Department of Health, Kinesiology, and Recreation, University of Utah College of Health, Salt Lake City, UT, United States

<sup>2</sup>Department of Geography, University of Utah, Salt Lake City, UT, United States

<sup>3</sup>School of Computing, University of Utah, Salt Lake City, UT, United States

<sup>4</sup>Department of Global Health, University of Washington, Seattle, WA, United States

<sup>5</sup>Department of Sociology, University of Utah, Salt Lake City, UT, United States

**Corresponding Author:**

Quynh C Nguyen, PhD

Department of Health, Kinesiology, and Recreation

University of Utah College of Health

1901 E. South Campus Drive

Annex Room 2124

Salt Lake City, UT,

United States

Phone: 1 801 585 5134

Fax: 1 801 585 3646

Email: [quynh.ctn@gmail.com](mailto:quynh.ctn@gmail.com)

## Abstract

**Background:** Studies suggest that where people live, play, and work can influence health and well-being. However, the dearth of neighborhood data, especially data that is timely and consistent across geographies, hinders understanding of the effects of neighborhoods on health. Social media data represents a possible new data resource for neighborhood research.

**Objective:** The aim of this study was to build, from geotagged Twitter data, a national neighborhood database with area-level indicators of well-being and health behaviors.

**Methods:** We utilized Twitter's streaming application programming interface to continuously collect a random 1% subset of publicly available geolocated tweets for 1 year (April 2015 to March 2016). We collected 80 million geotagged tweets from 603,363 unique Twitter users across the contiguous United States. We validated our machine learning algorithms for constructing indicators of happiness, food, and physical activity by comparing predicted values to those generated by human labelers. Geotagged tweets were spatially mapped to the 2010 census tract and zip code areas they fall within, which enabled further assessment of the associations between Twitter-derived neighborhood variables and neighborhood demographic, economic, business, and health characteristics.

**Results:** Machine labeled and manually labeled tweets had a high level of accuracy: 78% for happiness, 83% for food, and 85% for physical activity for dichotomized labels with the *F* scores 0.54, 0.86, and 0.90, respectively. About 20% of tweets were classified as happy. Relatively few terms (less than 25) were necessary to characterize the majority of tweets on food and physical activity. Data from over 70,000 census tracts from the United States suggest that census tract factors like percentage African American and economic disadvantage were associated with lower census tract happiness. Urbanicity was related to higher frequency of fast food tweets. Greater numbers of fast food restaurants predicted higher frequency of fast food mentions. Surprisingly, fitness centers and nature parks were only modestly associated with higher frequency of physical activity tweets. Greater state-level happiness, positivity toward physical activity, and positivity toward healthy foods, assessed via tweets, were associated with lower all-cause mortality and prevalence of chronic conditions such as obesity and diabetes and lower physical inactivity and smoking, controlling for state median income, median age, and percentage white non-Hispanic.

**Conclusions:** Machine learning algorithms can be built with relatively high accuracy to characterize sentiment, food, and physical activity mentions on social media. Such data can be utilized to construct neighborhood indicators consistently and cost effectively. Access to neighborhood data, in turn, can be leveraged to better understand neighborhood effects and address social



determinants of health. We found that neighborhoods with social and economic disadvantage, high urbanicity, and more fast food restaurants may exhibit lower happiness and fewer healthy behaviors.

(*JMIR Public Health Surveill* 2016;2(2):e158) doi:[10.2196/publichealth.5869](https://doi.org/10.2196/publichealth.5869)

## KEYWORDS

social media; Twitter messaging; health behavior; happiness; food; physical activity

## Introduction

There is an increasing recognition that health is determined by a myriad of factors, including where you live, play, and work [1-5]. Poor access to healthy food [6-10], abundance of fast food chains [11], lack of recreational facilities [12,13], and higher crime rates [7,14] have been shown to predict higher obesity rates. Environmental exposure to toxins, noise, and violence can be detrimental to health [15,16]. Conversely, neighborhood resources such as playgrounds for children, grocery stores, and gyms can be beneficial to health [17]. Adverse neighborhood conditions converge in poor, minority neighborhoods [18-21], thereby increasing health disparities.

Social environments can offer social and emotional support that buffers stressful life events [22]. Johns and colleagues found that neighborhoods with higher social cohesion had lower posttraumatic stress disorder [23]. Higher community happiness levels are linked with lower obesity, hypertension, and suicide rates as well as increased life expectancy [24-29]. Evidence also suggests that emotional states such as happiness, optimism, depression, or suicidality can spread through social networks [30-33]. The social environment can offer opportunities for social control in regulating unhealthy behaviors and facilitating the social learning of healthy behaviors but can also promote risky behaviors. Health behaviors, such as food consumption, health screening, smoking, alcohol consumption, drug use, and sleep have also been observed to spread through social networks [34-37].

The extreme scarcity of neighborhood data greatly limits research on neighborhood effects. Some places [38,39] have extensive neighborhood data collected on them, but they are the anomaly rather than the rule, and it is difficult to make comparisons across geographies because available measures vary greatly across them. Neighborhood data collection is expensive and time consuming and only available for certain time periods [40]. Widespread usage of the Internet and open recording of many transactions (eg, Yelp reviews, Foursquare check-ins, and reporting of personal opinions and behaviors through social media) has led to the availability of massive amounts of data that enable understanding of previously hidden local area interactions. Researchers are increasingly utilizing social media and user-generated data to track health behaviors and perform health surveillance (eg, for outbreak detection) [41-45]. Others have used social media to track sleep issues [46], personal health status disclosed by Twitter users [47,48], and patient-perceived quality of care [49].

In this study, we explored the utility of building a national neighborhood database from geotagged Twitter data to characterize well-being and health behaviors. We validated our

machine learning algorithm for constructing indicators of happiness, food, and physical activity by comparing machine-generated values to values generated by human labelers. In addition, we explored associations between Twitter-derived neighborhood variables and neighborhood demographic and economic characteristics. This project makes significant, relevant contributions to the field because neighborhood environments are increasingly linked to an array of important health outcomes and this project addresses the limits to research resulting from the lack of neighborhood data by providing new, cost-efficient data resources and methods for characterizing neighborhoods. To our knowledge, our study was the first to attempt to create a national neighborhood database from Twitter data, with indicators constructed for public health researchers. The only other type of neighborhood data that is consistently available for local areas is census data on the compositional characteristics of neighborhoods. Twitter is uniquely suited to characterize the social environment, including prevalent sentiment and health behaviors.

## Methods

### Social Media Data Collection

From February 2015 to March 2016, we utilized Twitter's streaming application programming interface (API) to continuously collect a random 1% sample of publicly available tweets with latitude and longitude coordinates. Given that neighborhood researchers differ in their use and interest in data at the census tract and zip code level, we constructed neighborhood indicators at both levels thereby increasing the flexibility of our dataset to address the potential data needs of other researchers. In total, we collected 79,848,992 million geotagged tweets from 603,363 unique Twitter users in the contiguous United States (including District of Columbia). The median number of tweets per user was 4. Job postings (identified through hashtags #hiring, #jobs, and #job) were removed from the final analytic sample of tweets because these were pervasive and not central to the neighborhood variables we constructed.

### Spatial Join and Neighborhood Definition

Each geotagged tweet was assigned a corresponding census tract and zip code it falls within, based on the latitude and longitude coordinates of where the tweet was sent. This spatial join procedure was implemented in Python (version 2.7.12; Python Software Foundation), a popular programming language for spatial data processing [50]. Specifically, Python libraries were used to read shapefile format vector data (PyShp 1.1.4), build an R-tree index on the polygon data (Rtree 0.8.2), and perform a spatial join operation (Shapely 1.5.12 and Fiona 1.6.1). The R-Tree was used to build a spatial index [51] on the national census tract and zip code polygon data to speed

computation. Tweets that were not assigned a census tract or zip code location included those with destinations bordering the United States (ie, Mexico and Canada). We linked 99.8% of tweets with geocoordinates to their respective 2010 census tract and zip code locations. The term *neighborhood* used in this paper refers to both zip codes and census tracts. We mapped tweets to these two geographic boundaries because they are among the most popular neighborhood definitions utilized by public health researchers [52-54].

### Processing Tweets

Duplicate tweets (ie, tweets with the same tweet ID, <1%) were removed computationally. Although Twitter's API collects a random subset of 1% of publicly available tweets, users (especially spam accounts) who tweet often have potentially greater influence on variable values we construct. We examined outliers in our datasets (defined as the users whose tweets accounted for more than 1% of tweets in our dataset) and eliminated automated accounts and accounts for which the majority of tweets were advertisements. Processing and statistical analysis tasks were performed with Stata MP13 (StataCorp LP).

### Construction of Neighborhood Variables From Twitter Data

From geotagged tweets, we derived variables that characterize happiness, food, and physical activity. Each tweet was divided into tokens using the Stanford tokenizer [55]. For processing of English text, tokens roughly correspond to words. We then built various algorithms utilizing tokens to create variables that characterize happiness and make references to food and physical activity. Below we describe in more detail our algorithms.

### Sentiment Analysis

To conduct sentiment analysis, we utilized the Machine Learning for Language Toolkit (MALLET; AK McCallum, 2002), a Java-based package for statistical natural language processing, document classification, clustering, topic modeling, information extraction, and other machine learning applications to text. We leveraged the Maximum Entropy text classifier in MALLET to classify tweets as happy and not happy [56]. In order to train our classifier, we obtained training sets from the following resources: Sentiment140 [57], Sanders Analytics [58], and Kaggle [59]. We trained our classifier to differentiate between happy and not happy sentiments. We then ran our classifier on our national Twitter data to compute a happy score (range 0-1) for each tweet, where higher happiness scores indicate more positive sentiment. MALLET estimates predicted probabilities that a tweet is happy based upon word-level features. The classifier uses search-based optimization to assign weights that maximize the likelihood of the training data. However, unlike Naïve Bayes, the Maximum Entropy classifier does not assume conditional independence among features.

To calibrate the generated happiness scores with human generated labels, two raters manually read a random subset of 1200 tweets and assigned a value of 1 to happy tweets and 0 to not happy tweets. The initial interrater reliability was 92%, and discordant values were reviewed until a 100% agreement between raters was reached. To decide on a cut point for

MALLET scores at which we would classify tweets as happy, we computed accuracy levels at different cut points of MALLET scores (Multimedia Appendix 1). Increasing the MALLET score improves the accuracy against human annotations but also reduces the calculated prevalence of tweets deemed as happy. A MALLET score of 0.80 achieves the highest level of accuracy while still maintaining a prevalence of happy tweets of 19% (which approximates the prevalence obtained by human annotations). Area under the receiver operating characteristic curve is approximately 0.7 for all MALLET cut points between 60 and 85.

### Food Analysis

We compiled a list of over 1430 popular food words from the US Department of Agriculture's National Nutrient Database [60]. Each food item was associated with a measure of caloric density, operationalized as calories per 100 grams. Fruits, vegetables, nuts, and lean proteins (ie, fish, chicken, and turkey) were labeled as healthy foods (340 food terms in total). Fried foods were not considered healthy foods. Our food list also contained popular national fast food restaurants such as McDonald's and Kentucky Fried Chicken (captured via 154 food terms including popular variations of restaurant names) to enable quantification of fast food references. From April 2015 to March 2016, we collected and processed 4,041,521 geotagged food tweets. In the food dataset, the median number of tweets per user was 12 tweets.

To analyze food culture, each tweet was examined for words or phrases matching those on our list. Each food item on our list was described by one or two words. Our text-matching algorithm first searched over a tweet for matches to two-word foods (eg, orange chicken). It then searched over the remaining words for matches to one-word food terms (eg, taco). We computed caloric density by summing up all the foods mentioned in the tweet. We also created a count of healthy food references and fast food restaurant references for each tweet. Moreover, we leveraged our sentiment analysis to assess sentiment toward food. Specifically, we tracked sentiment around healthy foods and fast food. These variables (any food references, healthy food references, fast food references, caloric density, and sentiment toward healthy foods and fast food) were then aggregated and summarized at the census tract and zip code level to create neighborhood indicators of food culture.

### Physical Activity Analysis

We created a list of physical activities using published lists of physical activity terms gathered from physical activity questionnaires, compendia of physical activities, and popularly available fitness programs [61,62]. Our physical activity list had 376 different activities that incorporate gym-related exercise (eg, treadmill, weight lifting), sports (eg, baseball), recreation (eg, hiking, scuba diving) and household chores (eg, gardening). We excluded popular phrases that generally do not relate to physical activity such as "walk away" and "running late." Using metabolic equivalents associated with physical activities, we quantified the exercise intensity of each physical activity mention, scaled for a duration of 30 minutes and for a 155-pound individual [63], which approximates the weight of an average American adult [64,65].

Upon piloting our algorithm, we identified commonly used phrases or pop culture references that do not involve physical activity (eg, walking dead) which were manually coded and excluded. Moreover, in order to help reduce the possibility that the tweet was about watching rather than actually participating in the physical activity, we excluded the tweet if it contained any of the following terms: “watch,” “watching,” “watches,” “watched,” “attend,” “attending,” “attends,” and “attended.” In reviewing preliminary labeled physical activity data, we found that most tweets (over 90%) pertaining to team sports (eg, baseball, basketball, football, soccer) were about watching games rather than participating in them. Thus, for team sports, we required that the tweet include the words “play,” “playing,” or “played.”

Our algorithm created the following physical activity variables for each tweet: any physical activity mention, exercise intensity, and sentiment around physical activity. From April 2015 to March 2016, we collected 1,473,976 geotagged physical activity tweets. In the physical activity dataset, the median number of tweets per user was 5 tweets.

### Quality Control Activities

A total of 5000 tweets have been manually labeled by two of the authors for quality control activities on food and physical activity. The authors manually labeled whether each tweet was food-related (2000), non-food-related (500), physical activity-related (2000), or non-physical activity-related (500). Excellent interrater reliability was achieved with greater than 90% agreement in all categories, and differences were discussed and resolved.

Among tweets our algorithm had labeled as food-related, 83% were labeled accurately when compared to labels generated by manual categorization. Among tweets our algorithm had labeled as non-food-related, 81% were labeled accurately (ie, both algorithm and human categorizers labeled the tweet as non-food-related). Overall, accuracy for food tweets was 83% and the *F* score was 0.86. It should be noted our algorithm could label a food-related tweet as non-food-related if the food reference was not in our food dictionary. Food items that are often associated with non-related food meaning, such as “perch,” have been excluded from our food dictionary. For tweets that had been mislabeled as food-related, common reasons included food term used as a metaphor, in a pun, or for food advertisement.

Among tweets our algorithm had labeled as physical activity-related, 82% of them were labeled accurately when compared to labels generated by human categorizers. An accuracy of 97% was found among tweets labeled as non-physical activity-related by our algorithm. The *F* score was 0.90 and the overall accuracy was 85% for physical activity tweets. Typical errors in classification of physical activity tweets included the use of an idiom (eg, running late) or the tweet was about watching sports games rather than playing sports.

Additionally, we evaluated our algorithm on its ability to identify relevant food and physical activity terms within tweets. To do this, we examined a random subset of tweets that the algorithm had identified as positive for food ( $n=200$ ) and

physical activity ( $n=200$ ). Here we focused on the accuracy of our algorithm to conduct string detection. We manually read the tweets to verify that manual annotations agreed with the terms detected. For food tweets, 87% of manual annotations matched all detected terms from the algorithm. Errors for nondetection of terms occurred when the tweet included a hashtag that had multiple food terms without spacing (eg, #chocolatebrownie) or when there were misspellings (eg, sandwich) or when the food was not on the food list. String detection for physical activity-related terms was more accurate with 98% of manual annotations matching detected terms from the algorithm. Errors included the omission of certain terms from the dictionary (eg, cycling) and use of hashtags without spacing of terms (#runrunrun).

We further evaluated our sentiment analysis activities through Amazon Mechanical Turk (Mturk; Amazon.com Inc, Seattle, WA, USA), an online crowdsourcing marketplace [66]. We randomly selected 500 tweets with 50% labeled as happy and 50% as not happy by our algorithm. Then, we created 20 online surveys through random sorting, with each survey consisting of 25 tweets. We asked participants to rate the sentiment of each tweet. All 20 surveys were live on April 1, 2015. Each online survey closed itself when 15 responses had been reached; the last survey closed on April 5, 2015. For each completed survey, 25 cents (\$0.25) was deposited into the participant’s Mturk account. A total of 32 participants completed 300 surveys (ie, 15 responses per survey, 20 surveys). Some participants completed multiple surveys rather than just one. Each tweet was then assigned a label of either happy or not happy based on the modal response from Mturkers (participants from Amazon Mturk). We found an accuracy of 69% for happy tweets and 80% for nonhappy tweets when compared to responses from Mturkers. The overall accuracy for sentiment was 78%, with an *F* score of 0.54.

We additionally compared performance of MALLETT with two other sentiment analysis techniques: a popular bag-of-words technique involving the use of a 10,000 word list [67] and Sentiment140, a machine-learning classifier [68]. Among the 500 control tweets from our LabMT experiment, the bag-of-words algorithm had an accuracy of 73% (*F* score 0.55) and Sentiment140 had an accuracy of 77% (*F* score 0.47).

### Other Publicly Available Neighborhood Data

To examine how Twitter-derived neighborhood variables relate to more traditional neighborhood variables, we merged our social media dataset with the 2010 Census and 2014 American Community Survey data which comprised the following demographic, household, and economic characteristics: household size, median family income and percent of the following: 65 years and older age group, 10-24 years, male, African American, white, Hispanic, households with relatives (other than spouse and children), households with unmarried partner, single female-headed households, householder living alone, owner-occupied housing, college graduates, unemployed, less than a high school degree and families living in poverty. A census tract was urban if the geographic centroid of the tract was in an area with more than 2500 people; all other tracts are rural. A zip code was defined as urban if the majority (75% or



more) of its land area was characterized as urban (ie, containing at least 2500 people).

Data on business types at the zip code level were obtained from the 2013 US Census Bureau zip code business patterns accessed via American FactFinder [69]. The following North American Industry Classification System (NAICS) codes were utilized to categorize businesses: 722410 (drinking places [alcoholic beverages]; these places are also known as bars, taverns, night clubs and primarily serve alcohol and may have limited food services) and 722511 (full-service restaurants; these include, for instance, diners and steakhouses). Fast food was defined by the following NAICS codes: 722513 (limited-service restaurants; these include carryout restaurants, drive-in restaurants, and other fast food restaurants) and 722515 (snack and nonalcoholic beverage bars). We also tracked supermarkets and grocery stores (NAICS code 445110) and convenience stores (NAICS code 445120). To examine associations between Twitter physical activity mentions and presence of recreational facilities, we retrieved business data for the following types of establishments: fitness and recreational sports centers (NAICS code 713940), nature parks (NAICS code 712190), zoos and botanical gardens (NAICS code 712130), golf courses and country clubs (NAICS code 713910), skiing facilities (NAICS code 713920), and bowling centers (NAICS code 713950).

We obtained state-level health outcome data including age-adjusted death rates due to all-causes and homicides from 2013 National Vital Statistics Reports. Data in this report was based on information from all resident death certificates filed in the 50 states and the District of Columbia. Death certificates are generally completed by funeral directors, attending physicians, medical examiners, and coroners. Age-adjusted death rates expressed per 100,000 population were based on the 2000 US standard population. Causes of death statistics were classified by the International Classification of Diseases, Tenth Revision, and based on the underlying cause of death.

We obtained age-adjusted prevalences of health risk behaviors and chronic conditions of US adult residents for the 50 states from the 2013 Behavioral Risk Factor Surveillance System (BRFSS), the nation's premier system of health-related telephone surveys. The questionnaires were created by BRFSS state coordinators and Centers for Disease Control and Prevention staff. BRFSS data includes self-reported physical activity, self-rated health, body mass index (BMI,  $\text{kg}/\text{m}^2$ ), and medical diagnoses of chronic conditions aggregated to the state level. Data from a national health survey suggests that BMI estimates derived from self-reported height and weight were lower than those are derived from measured height and weight, although BMI differences were generally less than  $1.0 \text{ kg}/\text{m}^2$  across sex and age groups [70]. State-level BRFSS data is publicly available. Smaller area aggregations can require data use agreements. In addition to state-level BRFSS data, we also utilized restricted-access zip-code-level data from the

2009-2014 Utah BRFSS survey to examine zip-code-level health outcomes [71,72].

## Regression Analyses

We implemented adjusted linear regression models to examine associations between area-level Twitter characteristics and other area-level characteristics (demographics, business characteristics, and health outcomes). To facilitate interpretation of findings for different variables, we standardized all variables to have a mean of zero and standard deviation of one. We investigated spatial autocorrelation and found that Moran's I was highest for census tract Twitter happiness (0.12) and less than 0.04 for other Twitter tract and zip code summaries. To account for spatial autocorrelation of area-level values in linear regression analyses, we adjusted standard errors for clustering of census tract and zip code values within a county. Statistical analyses were implemented with Stata MP13 (StataCorp LP) and ArcGIS Desktop version 10.1-10.3 (Esri).

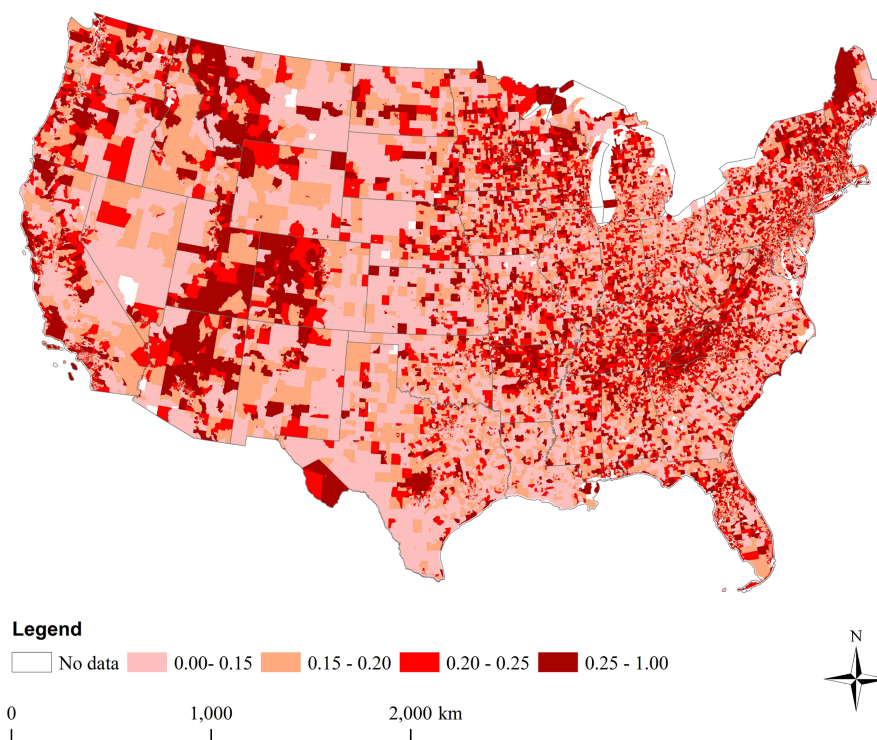
## Results

Table 1 displays descriptive statistics. Approximately 20% of tweets were happy. About 5.1% of tweets were about food and 1.8% were about physical activity. The mean and median caloric density of food references were 239 and 209 calories per 100 grams, respectively. Tweets about healthy food were happier than tweets about fast food (28.3% vs 14.5%;  $P < .001$ ). The mean and median exercise intensity of physical activity mentions (assuming 30 minutes for a 155-pound person) were 199 and 130 calories, respectively.

Figure 1 presents the spatial distribution of happy tweets by census tract, highlighting variation across the United States. Multimedia Appendix 2 presents the spatial distribution of happy tweets by zip code. The proportion of happy tweets was highest in the following states: Montana, Tennessee, Utah, New Hampshire, Arkansas, Maine, Colorado, and New York (Multimedia Appendix 3). By contrast, the proportions of happy tweets were lowest for the following states: Louisiana, North Dakota, Oregon, Maryland, Texas, Delaware, West Virginia, and Ohio (Multimedia Appendix 3).

Table 2 presents the results of adjusted linear regression analyses examining the associations between population characteristics and Twitter-derived characteristics at the census tract level (percent of tweets that were happy, percent of tweets about healthy food, percent of tweets about fast food, and percent of tweets about physical activity). Census tract characteristics like percent African American (beta coefficient,  $B = -.11$ ), greater household size ( $B = -.18$ ), and economic disadvantage ( $B = -.19$ ) were related to lower tract happiness. Economic disadvantage was negatively related to healthy food tweets ( $B = -.09$ ), fast food tweets ( $B = -.09$ ), and physical activity tweets ( $B = -.03$ ). Urbanicity was strongly related to higher frequency of fast food tweets ( $B = .29$ ). Greater household size was related to both lower healthy food tweets ( $B = -.11$ ) and fast food tweets ( $B = -.07$ ).

**Figure 1.** National distribution of happy tweets, by census tract. Geotagged tweets were spatially joined to their 2010 census tract locations and sentiment scores were computed. This color coded map presents the proportion of happy tweets in each census tract, with darker colors signifying higher proportions of happy tweets.



**Table 1.** Descriptive statistics of our national Twitter database, April 2015 to March 2016 (N=79,848,992).

	Mean (SD)
<b>Happiness</b>	
% Tweets that are happy	19.9 (6.7)
<b>Food culture</b>	
% Tweets about food	5.1 (22.0)
% Food tweets about healthy foods	15.9 (36.6)
% Food tweets about fast food	9.2 (29.0)
Caloric density of food tweets (per 100 grams)	238.5 (219.8)
% Food tweets that are happy	27.0 (44.4)
% Healthy food tweets that are happy	28.3 (45.0)
% Fast food tweets that are happy	14.5 (35.2)
<b>Physical activity culture</b>	
% Tweets about physical activity	1.8 (13.3)
Exercise intensity (per 30 minutes)	199.1 (117.5)
% Physical activity tweets that are happy	28.2 (45.0)



**Table 2.** Demographic and economic predictors of happy, food, and physical activity tweets from 70,515 census tracts (data source: 2010 US Census data).

Tract characteristics	% happy tweets	<i>P</i> value	% healthy food tweets	<i>P</i> value	% fast food tweets	<i>P</i> value	% physical activity tweets	<i>P</i> value
	Beta (95% CI) <sup>a</sup>		Beta (95% CI) <sup>a</sup>		Beta (95% CI) <sup>a</sup>		Beta (95% CI) <sup>a</sup>	
Urban (yes)	-.01 (-.04 to .03)	.79	.01 (-.02 to .03)	.54	.29 (.26 to .31)	<.001	-.02 (-.03 to -.01)	<.001
Population density	.06 (.03 to .08)	<.001	.04 (.02 to .07)	.001	-.03 (-.03 to -.02)	<.001	.00 (-.01 to .00)	.82
% 65 years and older	.02 (-.01 to .04)	.09	-.03 (-.04 to -.02)	<.001	-.03 (-.04 to -.01)	<.001	.02 (.02 to .03)	<.001
% 10-24 years	-.02 (-.04 to .00)	.01	-.05 (-.05 to -.04)	<.001	.00 (-.01 to .01)	.49	.00 (-.01 to .00)	.14
% Male	.04 (.03 to .06)	<.001	.01 (.00 to .02)	.21	-.05 (-.06 to -.04)	<.001	.01 (.01 to .02)	<.001
% African American	-.11 (-.14 to -.07)	<.001	-.03 (-.04 to -.01)	<.001	-.03 (-.04 to -.02)	<.001	-.01 (-.02 to -.01)	<.001
% Hispanic	-.04 (-.08 to .00)	.05	.02 (.01 to .03)	.00	.07 (.05 to .09)	<.001	.00 (.00 to .00)	.77
Household size	-.18 (-.20 to -.15)	<.001	-.11 (-.12 to -.09)	<.001	-.07 (-.09 to -.05)	<.001	-.01 (-.01 to -.01)	<.001
Economic disadvantage <sup>b</sup>	-.19 (-.21 to -.16)	<.001	-.09 (-.10 to -.08)	<.001	-.09 (-.10 to -.07)	<.001	-.03 (-.04 to -.03)	<.001

<sup>a</sup>Adjusted linear regression included all tract demographic and economic predictors simultaneously. Standard errors accounted for clustering at the county level.

<sup>b</sup>Economic disadvantage factor score derived from the following census tract characteristics: percent female-headed households, percent families living in poverty, unemployment rate, percent college graduates (reverse coded), and median family income (reverse coded).

Sensitivity analyses were performed to examine the relationship between population characteristics and happiness for a different unit of aggregation: zip code areas. Relationships seen at the census tract level were similar to those at the zip code level, although they were more muted at the zip code level (not shown). This may be the case because census tracts are designed to be relatively homogenous with regard to characteristics such as economic status and demographic characteristics [73].

Healthy foods (ie, vegetables, fruits, nuts, lean proteins) composed 15.9% of food tweets, while fast food restaurant mentions composed 9.2% of food tweets. The most popular foods include coffee, beer, pizza, wine, chicken, ice cream, and sushi (Figure 2). Popular healthy food terms included chicken, eggs, salad, turkey, and banana (Figure 3). Starbucks was the most popular fast food place mentioned (accounting for 46% of all fast food restaurant mentions), followed by Chipotle (9.2%), Taco Bell (5.4%), and Buffalo Wild Wings (5.2%). We additionally examined the relationship between food tweets and business characteristics. At the zip code level, greater numbers of fast food restaurants were associated with more fast food tweets ( $B=.15$ ), and higher caloric density of food mentions ( $B=.08$ ). Urban areas had tweets with higher caloric density ( $B=.08$ ) and more fast food restaurant mentions ( $B=.16$ ). Happy tweets were more prevalent in zip codes with higher numbers

of businesses ( $B=.11$ ) and full-service restaurants ( $B=.16$ ). Higher numbers of fast food restaurant ( $B=-.16$ ) and convenience stores ( $B=-.07$ ) were related to fewer happy tweets (Table 3).

Additionally, relatively few physical activity terms (13 terms) accounted for 75% of physical activity tweets (Figure 4) although our data collection system was set up to collect tweets on 376 physical activity terms. The most popular terms included walking, dancing, and running. At the zip code level, greater numbers of fitness and recreational sports centers were related to higher exercise intensity ( $B=.05$ ) and happier tweets ( $B=.07$ ). Surprisingly, the presence of nature parks was not associated with physical activity mentions. Urbanicity was associated with lower frequency of physical activity tweets and happy tweets but higher exercise intensity (Table 4). In supplemental analyses, we examined information on number of miles covered during physical activity if that was mentioned in the tweet ( $n=36,291$ ; median 3.1 miles). Even fewer tweets contained information on amount of time the person engaged in physical activity. Among 5823 tweets that mentioned hour(s) of physical activity, the median amount was 2 hours. Among 2402 tweets that only referred to minutes of physical activity, the median number of minutes was 20.

**Table 3.** Zip code and business characteristics as predictors of food tweets and happiness (data sources: 2013 zip code business patterns and 2010 US Census data).

Zip code characteristics	Average caloric density of food tweets n=21,756 Beta (95% CI) <sup>a</sup>	P value	% fast food tweets n=21,756 Beta (95% CI) <sup>a</sup>	P value	% happy tweets n=26,584 Beta (95% CI) <sup>a</sup>	P value
Urban (yes)	.08 (.05 to .11)	<.001	.16 (.12 to .20)	<.001	-.02 (-.06 to .02)	.29
Population density	.00 (.00 to .01)	.24	.00 (-.01 to .01)	.86	.01 (.00 to .03)	.18
Number of businesses	-.01 (-.02 to .01)	.34	.02 (.00 to .04)	.04	.11 (.08 to .15)	<.001
Businesses that sell alcohol	-.03 (-.04 to -.02)	<.001	-.04 (-.05 to -.04)	<.001	-.01 (-.02 to .00)	.02
Full service restaurants	-.04 (-.06 to -.02)	<.001	.01 (-.01 to .03)	.43	.16 (.13 to .20)	<.001
Fast food restaurants	.08 (.06 to .10)	<.001	.15 (.13 to .17)	<.001	-.16 (-.20 to -.12)	<.001
Grocery stores	.01 (.00 to .01)	.28	-.04 (-.05 to -.03)	<.001	-.02 (-.04 to .00)	.05
Convenience stores	.02 (.01 to .02)	<.001	-.03 (-.04 to -.02)	<.001	-.07 (-.08 to -.05)	<.001

<sup>a</sup>Adjusted linear regression included all zip code and business characteristics simultaneously. Standard errors accounted for clustering at the county level.

**Table 4.** Zip code and business characteristics as predictors of physical activity tweets and happiness (data sources: 2013 zip code business patterns and 2010 US Census data).

Zip code characteristics	% physical activity tweets n=26,839 Beta (95% CI) <sup>a</sup>	P value	Exercise intensity n=20,715 Beta (95% CI) <sup>a</sup>	P value	% happy tweets n=26,839 Beta (95% CI) <sup>a</sup>	P value
Urban (yes)	-.09 (-.11 to -.07)	<.001	.07 (.04 to .11)	<.001	-.08 (-.12 to -.04)	<.001
Population density	-.01 (-.02 to .00)	.01	-.01 (-.01 to .00)	.03	.01 (.00 to .02)	.08
Fitness/recreational centers	.01 (.00 to .02)	.003	.05 (.04 to .06)	<.001	.07 (.06 to .08)	<.001
Nature parks	.01 (.00 to .02)	.05	-.01 (-.01 to .00)	.21	.03 (.02 to .04)	<.001
Zoos/botanical gardens	.00 (.00 to .01)	.19	.00 (-.01 to .00)	.35	.02 (.01 to .03)	<.001
Golf/country clubs	.03 (.02 to .03)	<.001	-.05 (-.06 to -.04)	<.001	.03 (.02 to .04)	<.001
Skiing facilities	.04 (.04 to .05)	<.001	.02 (.02 to .03)	<.001	.03 (.02 to .03)	<.001
Bowling centers	-.01 (-.02 to -.01)	<.001	-.01 (-.02 to .00)	.01	-.02 (-.03 to -.01)	<.001

<sup>a</sup>Adjusted linear regression included all zip code and business characteristics simultaneously. Standard errors accounted for clustering at county level.

**Table 5.** Twitter happiness as a predictor of health outcomes in 232 zip codes in Utah (data source: Utah Behavioral Risk Factor Surveillance System [BRFSS] survey 2009-2014. BRFSS underwent design feature changes. Life dissatisfaction values were only available for 2009 and 2010. All other variables were averages from available data from 2011-2014).

Zip code health outcomes	Beta (95% CI) <sup>a</sup> n=232	P value
Life dissatisfaction	.01 (-.13 to .15)	.91
Self-rated health (higher score=worse health)	-.08 (-.21 to .05)	.21
Any past month physical activity/exercise	.13 (.00 to .26)	.05
Body mass index (kg/m <sup>2</sup> )	-.13 (-.26 to -.01)	.04

<sup>a</sup>Separate linear regression models for each zip code health outcome.

**Table 6.** State level Twitter sentiment predictors of health outcomes (N=49 states in the contiguous United States plus District of Columbia. Data sources: 2013 National Vital Statistics Reports and 2013 Behavioral Risk Factor Surveillance System [BRFSS] survey on adults).

State-level adult health outcomes	Twitter predictor variables					
	Happiness Beta (95% CI) <sup>a</sup>	P value	Positive sentiment toward healthy foods Beta (95% CI) <sup>a</sup>	P value	Positive sentiment toward physical activity Beta (95% CI) <sup>a</sup>	P value
All-cause mortality per 100,000	-32.34 (-61.59 to -3.09)	.03	-23.51 (-40.54 to -6.48)	.01	-25.37 (-42.00 to -8.74)	.004
Homicide per 100,000	-1.02 (-1.98 to -.06)	.03	-.76 (-1.28 to -.25)	.01	-.75 (-1.28 to -.23)	.01
% With diabetes	-.58 (-1.05 to -.12)	.02	-.52 (-.78 to -.27)	<.001	-.41 (-.68 to -.14)	.004
% With obesity	-2.27 (-3.35 to -1.18)	<.001	-1.67 (-2.25 to -1.09)	<.001	-1.43 (-2.05 to -.80)	<.001
% Poor/fair self-rated health	-1.13 (-2.13 to -.13)	.03	-.77 (-1.36 to -.19)	.01	-.61 (-1.21 to -.02)	.05
% With high cholesterol	-.78 (-1.66 to .11)	.08	-.51 (-1.04 to .01)	.06	-.75 (-1.25 to -.26)	.003
% Physical inactivity	-2.46 (-4.80 to -.12)	.04	-2.32 (-3.61 to -1.03)	.001	-1.59 (-2.97 to -.22)	.02
% Current smoking	-1.47 (-2.68 to -.27)	.02	-1.20 (-1.88 to -.52)	.001	-1.14 (-1.82 to -.45)	.002

<sup>a</sup>Each cell in the table represents the coefficient estimate of the predictor variable (given by the column) on the state-level health outcome (given by the row). Adjusted linear regression models controlled for state-level demographics: median age, % non-Hispanic white, median household income.

Additionally, merging in health-related datasets, we examined associations between our Twitter-based variables and other measures of health and well-being. Utilizing data from the 2009-2014 BRFSS in Utah, we found that zip codes in Utah with higher Twitter happiness scores were associated with lower body mass index and higher physical activity (Table 5). However, Twitter happiness scores were not statistically significantly related to self-rated health or life satisfaction.

Greater state-level happiness, as indicated by tweets, was related to lower prevalence of obesity; a one standard deviation increase in happiness was associated with two percentage points lower prevalence in obesity. Greater positive sentiment for healthy foods was related to lower prevalence of diabetes and obesity

and lower percent of the population who are physically inactive or current smokers (Table 6). Positive sentiment toward physical activity was related to lower obesity.

Table 7 presents adjusted regression results for additional Twitter-derived variables (percentage of food tweets about healthy foods, percentage of food tweets about fast food, and percentage of tweets about physical activity) and a select number of state health outcomes. Out of the three Twitter-derived variables, percentage of tweets about physical activity was the strongest and most consistent predictor; more online discussion about physical activity was related to lower all-cause mortality and lower prevalence of obesity and fair/poor self-rated health.

**Table 7.** State level Twitter food and physical activity characteristics as predictors of health outcomes (N=49 states in the contiguous United States plus District of Columbia. Data sources: 2013 National Vital Statistics Reports and 2013 Behavioral Risk Factor Surveillance System [BRFSS] survey on adults).

Twitter predictors	State-level adult health outcomes					
	All-cause mortality per 100,000 Beta (95% CI) <sup>a</sup>	P value	% with obesity Beta (95% CI) <sup>a</sup>	P value	% poor/fair self-rated health Beta (95% CI) <sup>a</sup>	P value
% Of food tweets about healthy food	11.74 (-6.48 to 29.96)	.20	-.09 (-.64 to .45)	.73	.11 (-.48 to .70)	.71
% Of food tweets about fast food	9.84 (-8.56 to 28.25)	.29	.68 (.13 to 1.23)	.02	.77 (.18 to 1.37)	.01
% Of tweets about physical activity	-28.17 (-46.68 to -9.65)	.004	-1.86 (-2.41 to -1.31)	<.001	-.89 (-1.49 to -.29)	.01

<sup>a</sup>Adjusted linear regression models were run separately for each state-level health outcome (column) and included all three predictors (row) simultaneously in addition to the following state-level control variables: median age, % non-Hispanic white, median household income. Beta coefficient represents a change in the outcome for every standard deviation change in the predictor (row variable).

Figure 2. Items in the top 50% of food tweets.

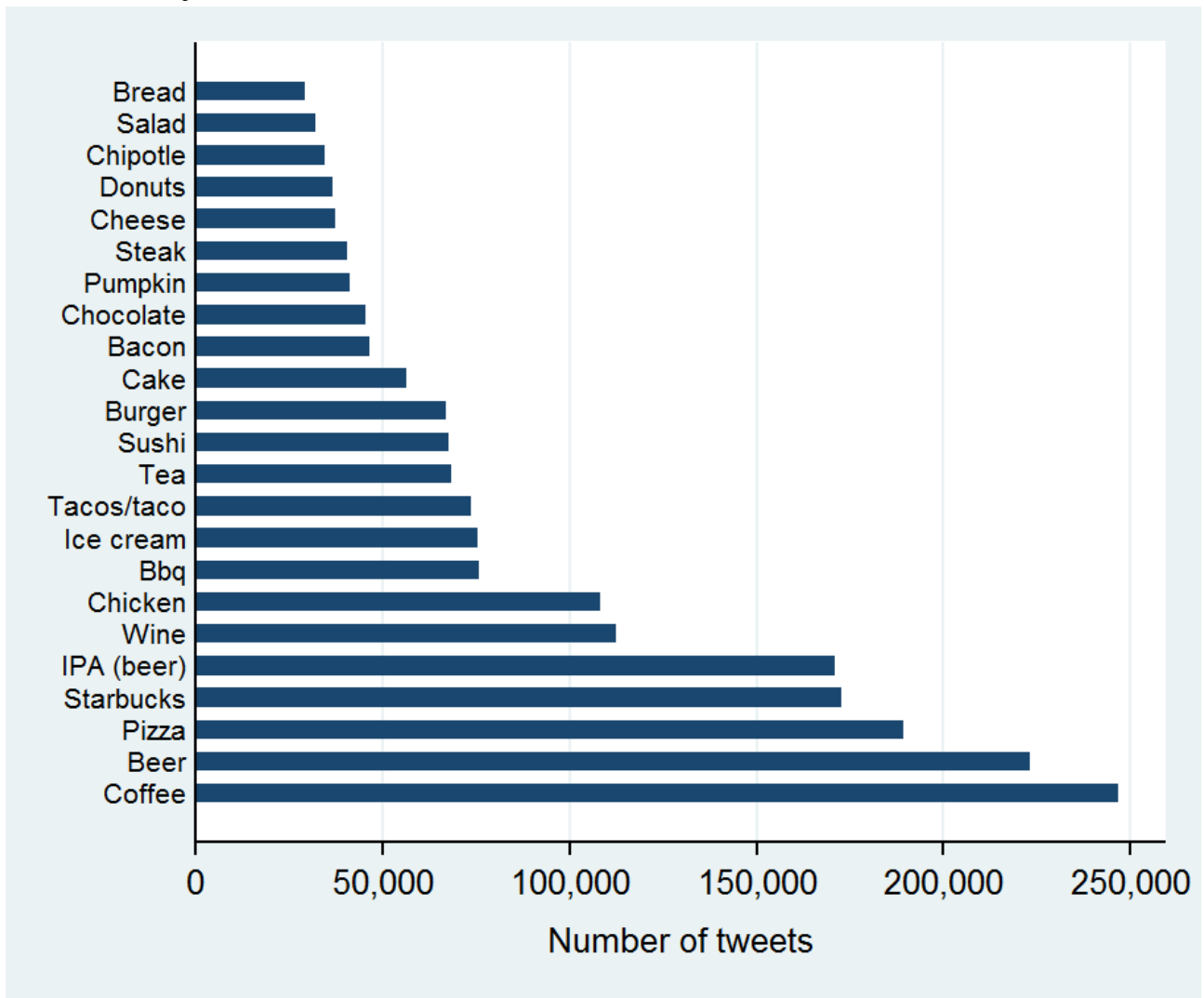
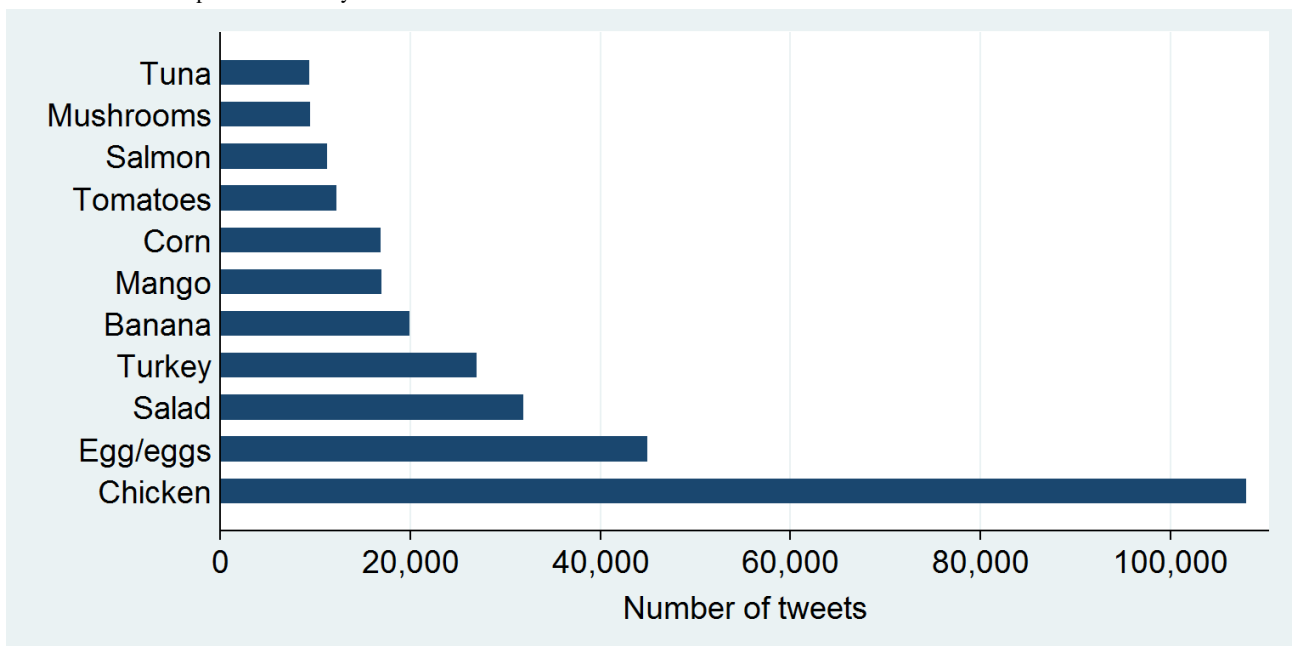
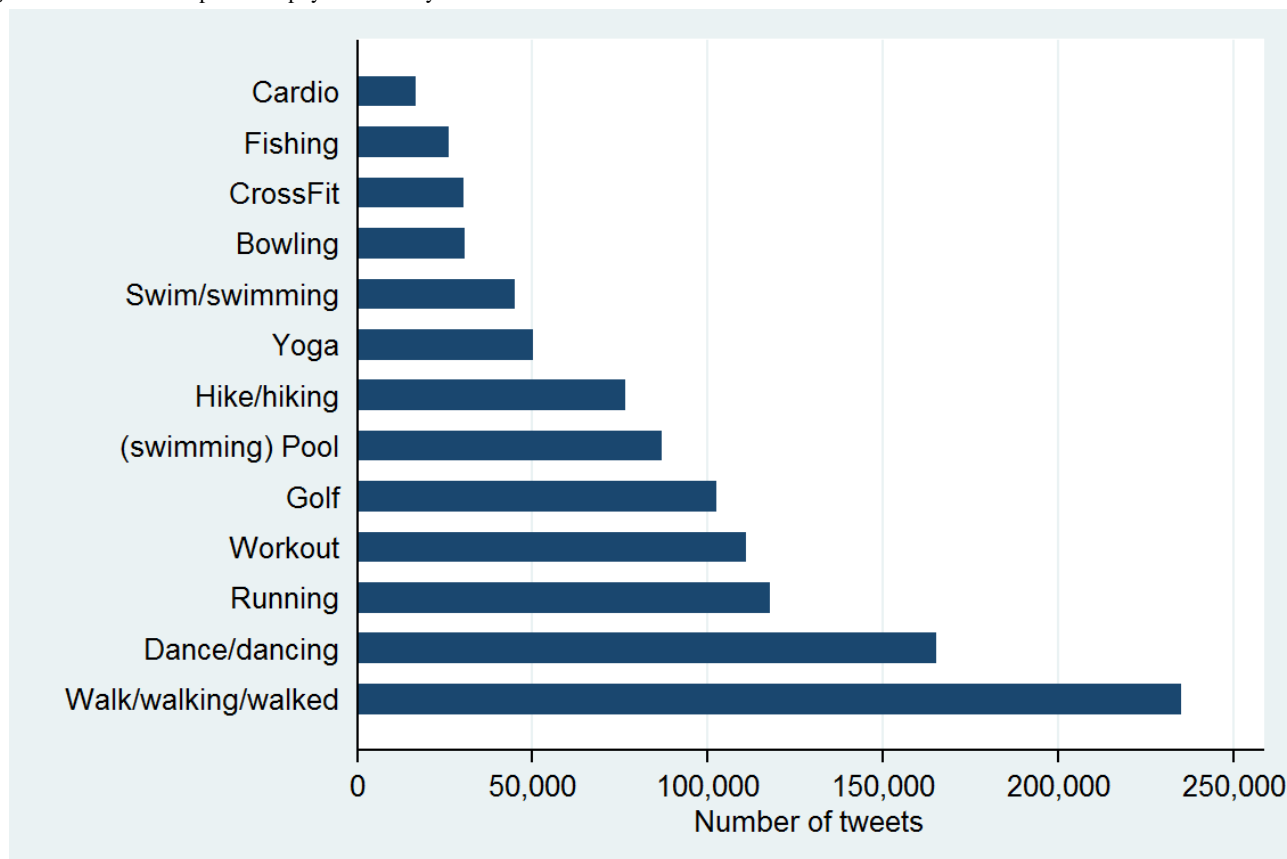


Figure 3. Items in the top 50% of healthy food tweets.



**Figure 4.** Items in the top 75% of physical activity tweets.

## Discussion

### Principal Findings

In this paper, we detail the building of a new national neighborhood data repository constructed from Twitter data which addresses a pressing need for neighborhood data that are available across large geographies and can be updated efficiently and cost-effectively. We demonstrate that simple machine learning algorithms for the construction of indicators for happiness, food, and physical activity can agree extremely well with manually generated labels. About one-fifth of tweets were identified as happy. There was substantial spatial variation in happiness across the United States. For instance, the proportion of tweets that were happy in Montana (the most happy state) was 10% greater than in Louisiana (the least happy state). Only a few terms are needed to capture the majority of tweets on food and physical activity. Economic disadvantage, urbanicity, and presence of fast food restaurants predicted lower area level happiness and lower frequency of healthy behavior mentions on Twitter. Moreover, we find that Twitter area-level characteristics are correlated with area-level health outcomes relating to health behaviors, chronic diseases, mortality, and self-rated health.

### Study Findings in Context

Social media represents an important new data resource that is increasingly being harnessed for public health efforts such as surveillance of smoking behavior and sentiment toward tobacco products [74]. However, few studies are leveraging social media data for the investigation of local area characteristics. More

commonly, studies utilizing social media data examine patterns at the city, county, or state level [67,75] rather than at finer levels of aggregation, which is necessary for understanding the potential impacts of neighborhood conditions.

Neighborhoods can impact health through a myriad of pathways. Disadvantaged neighborhoods may have fewer resources that support physical activity and healthy diets. Poor and minority neighborhoods have fewer large supermarkets (where healthy foods are more abundant and affordable) compared to wealthy and majority white neighborhoods. Studies have documented increased fruit and vegetable consumption with more supermarket availability [17]. Poor neighborhoods, which have been labeled food deserts, also tend to have more fast food restaurants, which can contribute to weight gain [6]. In this study, we found that higher numbers of fast food restaurants were associated with higher frequency of fast food mentions, lower frequency of healthy food mentions, and less positive sentiment about healthy foods on Twitter. Our results align with a recent study conducted analyzing Instagram posts, which found that posts originating from census tracts deemed as food deserts contained fewer mentions of fruits and vegetables compared to Instagram posts outside food deserts [76]. Additionally, neighborhoods may promote poor health through psychosocial pathways. Living in neighborhoods that are unclean, noisy, and violent can be psychologically harmful through over-activation of the stress response [77,78].

We found that economic disadvantage was related to lower frequency of happy tweets. Previous research by Mitchell and colleagues found that higher socioeconomic status was associated with higher Twitter happiness scores at the city level.



Moreover, they identified mild correlations ( $r=-0.34$ ) between happiness and obesity rates for 190 metropolitan statistical areas [67] and that Twitter happiness scores were moderately correlated with other state-level indicators of well-being including shootings, the Peace index, America's Health Ranking, and the Gallup-Healthways Well-Being Index (correlations ranged between 0.51 and 0.64) [67].

### Study Strengths and Limitations

In this paper, we describe the creation of a new neighborhood data repository constructed from Twitter data and merged with publicly available administrative datasets. However, this study is subject to several limitations. For instance, users of social media tend to be younger; in 2014, 37% of individuals aged 18 to 29 years old used Twitter compared to 12% of individuals aged 50 to 64 years and 10% among those 65 years and older. Nonetheless, adoption rates of social media have been steadily increasing [79]. Tweets also include information rarely found in other neighborhood sources. Twitter users are composed of individuals as well as groups of individuals, organizations, companies, and news outlets. Thus, compiling such information may allow for a more comprehensive examination of the social environment.

Moreover, we are only collecting a subset of publicly available tweets, and thus conclusions from our analytic sample may not generalize to the full population of tweets [80]. Our construction of neighborhood indicators from Twitter data necessitated that we restricted our data collection to geolocated tweets. We utilized Twitter's API which allows the retrieval of a maximum resulting volume of 1% of the total tweets at any given time point. Previous studies suggest that about 1% to 2% of tweets may contain global positioning system location information [81,82] and that use of Twitter's streaming API may obtain 40% to 90% of all geotagged tweets [81,82]. Tweets with location information may be different from those without. For example, tweets in which users share their locations may be more likely to contain public and social activities such as friends tweeting from a restaurant or an event. However, in sensitivity analyses with a subset of control tweets ( $n=138,152$  tweets) collected from July 9 to July 14, 2015, we did not detect any statistically significant differences in happiness scores between tweets with and without geographic coordinates (not shown).

In creating our neighborhood indicators from Twitter data, we prioritized transparency and ease of implementation so that other researchers can replicate our algorithms. Our sentiment algorithm was trained to differentiate between happy and not happy sentiments (which encompasses neutral and sad

sentiments). Thus, we were not able to specifically examine the prevalence of sad tweets, which may provide additionally useful information about the well-being of communities. In future work, we plan to target the identification of sadness. Our algorithms for food and physical activity implemented corpus-based classification with steps that are easily understandable. However, this technique does not take into account the entire context of sarcasm or humor in a tweet, challenges which still evade most natural language processing algorithms though some studies show promising results [83,84]. Our analysis of caloric density of food assumed calories per 100 grams. Most tweets do not specify the exact amount of food consumed, and thus our estimate is just an approximation.

Additionally, the content of tweets reflects the type of information that people feel comfortable reporting and may not represent the true spectrum of their feelings or their experiences. For instance, people may feel most comfortable presenting a neutral stance rather than voicing polarizing viewpoints. Certain foods (cupcakes) may get tweeted more often than others (celery). Additionally, we cannot be certain that the food that was tweeted was indeed consumed. Similarly, physical activity tweets may reflect a mixture of intentions, plans, and actual engagement in those physical activities. Also, exercise intensity for physical activities was assessed for 30 minutes of physical activity for an individual weighing 155 pounds, which can be an under- or overestimation depending on the type of activity and persons engaged in that activity.

### Conclusions

The epidemic rise in obesity and related chronic diseases in recent decades signal the importance of structural forces and social processes, but the dearth of data on contextual factors limits the investigation of multilevel effects on health. Social media data can be uniquely harnessed to capture social and cultural processes with potential impacts on health [71,72,85-89]. For instance, public posts can be utilized to measure prevalent happiness which can impact health through emotional contagion and the interconnectedness between mental health and physical health. Additionally, public posts about health behaviors may help us understand the prevalence of those behaviors as well as local area social norms. We demonstrate that tweets can provide a means to assess prevalent sentiment and food behaviors and physical activity, which can inform health interventions and policies to meet the needs of different neighborhoods. In particular, as this study suggests, neighborhoods with social and economic disadvantage, high urbanicity, and those with more fast food restaurants may exhibit lower happiness and fewer healthy behaviors.

### Acknowledgments

This work was supported by a National Institutes of Health grant (5K01ES025433) to Dr Nguyen. The research uses data from the Utah BRFSS survey, which is implemented by the Utah Department of Health in conjunction with the US Centers for Disease Control and Prevention. We thank Patsaporn Kanokvimankul for her assistance with locating some of the external health outcomes data for this paper. We thank Drs Jared B Hawkins and John S Brownstein for their assistance with quality control activities associated with the Twitter data.

## Conflicts of Interest

None declared.

## Multimedia Appendix 1

Varying MALLEET cut points for happy tweets and comparisons with manually generated labels.

[PDF File (Adobe PDF File), 14KB - [publichealth\\_v2i2e158\\_app1.pdf](#)]

## Multimedia Appendix 2

National distribution of happy tweets, by zip code. Geotagged tweets were spatially joined to their 2010 zip code locations and sentiment scores were computed. This color coded map presents the proportion of happy tweets in each zip code area, with darker colors signifying higher proportions of happy tweets.

[JPG File, 6MB - [publichealth\\_v2i2e158\\_app2.jpg](#)]

## Multimedia Appendix 3

Proportion of happy tweets, by state.

[PDF File (Adobe PDF File), 18KB - [publichealth\\_v2i2e158\\_app3.pdf](#)]

## References

1. Social determinants of health. Washington, DC: US Department of Health and Human Services; 2016 Jan 30. URL: <https://www.healthypeople.gov/2020/topics-objectives/topic/social-determinants-of-health> [accessed 2016-09-28] [WebCite Cache ID 6ks1I76Df]
2. Marmot M. Closing the gap in a generation: health equity through action on the social determinants of health. *Lancet* 2008 Nov 08;372(9650):1661-1669.
3. Villanueva K. People living in hilly residential areas in metropolitan Perth have less diabetes: spurious association or important environmental determinant? *Int J Health Geogr* 2013;12(1):1-11.
4. Schmidt NM, Lincoln AK, Nguyen QC, Acevedo-Garcia D, Osypuk TL. Examining mediators of housing mobility on adolescent asthma: results from a housing voucher experiment. *Soc Sci Med* 2014 Apr;107:136-144 [FREE Full text] [doi: [10.1016/j.socscimed.2014.02.020](https://doi.org/10.1016/j.socscimed.2014.02.020)] [Medline: [24607675](#)]
5. Nguyen QC, Rehkopf DH, Schmidt NM, Osypuk TL. Heterogeneous effects of housing vouchers on the mental health of US adolescents. *Am J Public Health* 2016 Apr;106(4):755-762. [doi: [10.2105/AJPH.2015.303006](https://doi.org/10.2105/AJPH.2015.303006)] [Medline: [26794179](#)]
6. 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](#)]
7. Stafford M, Cummins S, Ellaway A, Sacker A, Wiggins RD, Macintyre S. Pathways to obesity: identifying local, modifiable determinants of physical activity and diet. *Soc Sci Med* 2007 Nov;65(9):1882-1897. [doi: [10.1016/j.socscimed.2007.05.042](https://doi.org/10.1016/j.socscimed.2007.05.042)] [Medline: [17640787](#)]
8. Wang MC, Kim S, Gonzalez AA, MacLeod KE, Winkleby MA. Socioeconomic and food-related physical characteristics of the neighbourhood environment are associated with body mass index. *J Epidemiol Community Health* 2007 Jun;61(6):491-498 [FREE Full text] [doi: [10.1136/jech.2006.051680](https://doi.org/10.1136/jech.2006.051680)] [Medline: [17496257](#)]
9. Inagami S, Cohen DA, Finch BK, Asch SM. You are where you shop: grocery store locations, weight, and neighborhoods. *Am J Prev Med* 2006 Jul;31(1):10-17. [doi: [10.1016/j.amepre.2006.03.019](https://doi.org/10.1016/j.amepre.2006.03.019)] [Medline: [16777537](#)]
10. Christiansen KMH, Qureshi F, Schaible A, Park S, Gittelsohn J. Environmental factors that impact the eating behaviors of low-income African American adolescents in Baltimore City. *J Nutr Educ Behav* 2013;45(6):652-660. [doi: [10.1016/j.jneb.2013.05.009](https://doi.org/10.1016/j.jneb.2013.05.009)] [Medline: [23916684](#)]
11. Block JP, Scribner RA, DeSalvo KB. Fast food, race/ethnicity, and income: A geographic analysis. *Am J Prev Med* 2004;27(3):211-217.
12. Roemmich JN, Epstein LH, Raja S, Yin L, Robinson J, Winiewicz D. Association of access to parks and recreational facilities with the physical activity of young children. *Prev Med* 2006 Dec;43(6):437-441. [doi: [10.1016/j.ypmed.2006.07.007](https://doi.org/10.1016/j.ypmed.2006.07.007)] [Medline: [16928396](#)]
13. Brownson RC, Hoehner CM, Day K, Forsyth A, Sallis JF. Measuring the built environment for physical activity: state of the science. *Am J Prev Med* 2009 Apr;36(4 Suppl):S99-S123 [FREE Full text] [doi: [10.1016/j.amepre.2009.01.005](https://doi.org/10.1016/j.amepre.2009.01.005)] [Medline: [19285216](#)]
14. Mujahid MS, Diez Roux AV, Shen M, Gowda D, Sánchez B, Shea S, et al. Relation between neighborhood environments and obesity in the multi-ethnic study of atherosclerosis. *Am J Epidemiol* 2008 Jun 1;167(11):1349-1357 [FREE Full text] [doi: [10.1093/aje/kwn047](https://doi.org/10.1093/aje/kwn047)] [Medline: [18367469](#)]

15. Yen IH, Kaplan GA. Poverty area residence and changes in physical activity level: evidence from the Alameda County Study. *Am J Public Health* 1998;88(11):1709-1712.
16. Ross CE. Walking, exercising, and smoking: does neighborhood matter? *Soc Sci Med* 2000 Jul;51(2):265-274. [Medline: [10832573](#)]
17. Morland K, Wing S, Diez RA. The contextual effect of the local food environment on residents' diets: the atherosclerosis risk in communities study. *Am J Public Health* 2002 Nov;92(11):1761-1767. [Medline: [12406805](#)]
18. Black JL, Macinko J, Dixon LB, Fryer GE. Neighborhoods and obesity in New York City. *Health Place* 2010 May;16(3):489-499. [doi: [10.1016/j.healthplace.2009.12.007](#)] [Medline: [20106710](#)]
19. Diez-Roux AV. Bringing context back into epidemiology: variables and fallacies in multilevel analysis. *Am J Public Health* 1998 Feb;88(2):216-222. [Medline: [9491010](#)]
20. Macintyre S, MacIver S, Sooman A. Area, class and health: Should we be focusing on places or people. *J Soc Policy* 1993;22:213-233.
21. Duncan C, Jones K, Moon G. Context, composition and heterogeneity: using multilevel models in health research. *Soc Sci Med* 1998;46:97-117.
22. Pearlin LI. The sociological study of stress. *J Health Soc Behav* 1989;30(3):241-256.
23. Johns LE. Neighborhood social cohesion and posttraumatic stress disorder in a community-based sample: findings from the Detroit Neighborhood Health Study. *Soc Psych Psych Epid* 2012;47(12):1899-1906.
24. Oswald AJ, Powdthavee N. Obesity, unhappiness, and the challenge of affluence: theory and evidence. *Econ J* 2007;117:117.
25. Bray I, Gunnell D. Suicide rates, life satisfaction and happiness as markers for population mental health. *Soc Psych Psych Epid* 2006;41(5):333-337.
26. Tella RD, MacCulloch RJ, Oswald AJ. The macroeconomics of happiness. *Rev Econ Stat* 2003;85(4):809-827.
27. Blanchflower DG, Oswald AJ. Hypertension and happiness across nations. *J Health Econ* 2008;27(2):218-233.
28. Dodds PS. Temporal patterns of happiness: information in a global social network: hedonometrics and Twitter. *PLoS ONE* 2011;6(12):e26752.
29. Di Tella R, MacCulloch R. Gross national happiness as an answer to the Easterlin Paradox? *J Devel Econ* 2008;86(1):22-42.
30. Bearman PS, Moody J. Suicide and friendships among American adolescents. *Am J Public Health* 2004;94(1):89-95.
31. Larson R, Richards MH. *Divergent Realities: The Emotional Lives of Mothers, Fathers, and Adolescents*. New York, NY: Basic Books; 1994.
32. Fowler JH, Christakis N. Dynamic spread of happiness in a large social network: Longitudinal analysis over 20 years in the Framingham heart study. *Brit Med J* 2008;337:a2338.
33. Guan W, Kamo Y. Contextualizing depressive contagion: A multilevel network approach. *Soc Ment Health* 2015 Dec 09. [doi: [10.1177/2156869315619657](#)]
34. Pachucki MA, Jacques PF, Christakis NA. Social network concordance in food choice among spouses, friends, and siblings. *Am J Public Health* 2011 Nov;101(11):2170-2177 [FREE Full text] [doi: [10.2105/AJPH.2011.300282](#)] [Medline: [21940920](#)]
35. Keating NL, O'Malley AJ, Murabito JM, Smith KP, Christakis NA. Minimal social network effects evident in cancer screening behavior. *Cancer* 2011 Jul 1;117(13):3045-3052 [FREE Full text] [doi: [10.1002/ncr.25849](#)] [Medline: [21264828](#)]
36. Rosenquist JN, Murabito J, Fowler JH, Christakis NA. The spread of alcohol consumption behavior in a large social network. *Ann Intern Med* 2010 Apr 6;152(7):426-433 [FREE Full text] [doi: [10.7326/0003-4819-152-7-201004060-00007](#)] [Medline: [20368648](#)]
37. Mednick SC, Christakis NA, Fowler JH. The spread of sleep loss influences drug use in adolescent social networks. *PLoS One* 2010;5(3):e9775 [FREE Full text] [doi: [10.1371/journal.pone.0009775](#)] [Medline: [20333306](#)]
38. National Archive of Criminal Justice. 2012. Project on Human Development in Chicago Neighborhoods URL: <http://www.icpsr.umich.edu/icpsrweb/PHDCN/> [accessed 2016-09-28] [WebCite Cache ID 6ks3vmrzJ]
39. Baltimore Neighborhood Indicators Alliance: Vital Signs 11. 2013 Sep 24. URL: <http://bniajfi.org/wp-content/uploads/2014/04/Vs-11-Intro.pdf> [accessed 2016-09-28] [WebCite Cache ID 6ks49HpP9]
40. Peterson RD, Krivo LJ. National Neighborhood Crime Study. 2000. URL: <http://www.icpsr.umich.edu/icpsrweb/RCMD/studies/27501> [accessed 2016-09-28] [WebCite Cache ID 6ks4ICpcz]
41. Eysenbach G. Infodemiology: tracking flu-related searches on the web for syndromic surveillance. *AMIA Annu Symp Proc* 2006:244-248 [FREE Full text] [Medline: [17238340](#)]
42. Eysenbach G. Infodemiology and infoveillance. *Am J Prev Med* 2011;40(5):S154-S158.
43. Yepes AJ, Han B. Investigating public health surveillance using Twitter. *ACL-IJCNLP* 2015;2015:164.
44. Nsoesie EO, Klueberg 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](#)]
45. Woo H, Cho Y, Shim E, Lee J, Lee C, Kim SH. Estimating influenza outbreaks using both search engine query data and social media data in South Korea. *J Med Internet Res* 2016;18(7):e177 [FREE Full text] [doi: [10.2196/jmir.4955](#)] [Medline: [27377323](#)]
46. McIver DJ, Hawkins JB, Chunara R, Chatterjee AK, Bhandari A, Fitzgerald TP, et al. Characterizing sleep issues using Twitter. *J Med Internet Res* 2015;17(6):e140 [FREE Full text] [doi: [10.2196/jmir.4476](#)] [Medline: [26054530](#)]

47. Nguyen QC. Leveraging geotagged Twitter data to examine neighborhood happiness, diet, and physical activity. *Appl Geogr* 2016;73:77-88.
48. Yin Z, Fabbri D, Rosenbloom ST, Malin B. A scalable framework to detect personal health mentions on Twitter. *J Med Internet Res* 2015;17(6):e138 [FREE Full text] [doi: [10.2196/jmir.4305](https://doi.org/10.2196/jmir.4305)] [Medline: [26048075](https://pubmed.ncbi.nlm.nih.gov/26048075/)]
49. Hawkins JB. Measuring patient-perceived quality of care in US hospitals using Twitter. *BMJ Qual Safety* 2015:4309.
50. Etherington TR. Teaching introductory GIS programming to geographers using an open source Python approach. *J Geogr Higher Educ* 2016;40(1):117-130.
51. Guttman A. R-trees: a dynamic index structure for spatial searching. 1984 Presented at: 1984 ACM SIGMOD international conference on Management of Data; 1984; New York, NY p. 47-57.
52. Nguyen QC, Schmidt NM, Glymour MM, Rehkopf DH, Osypuk TL. Were the mental health benefits of a housing mobility intervention larger for adolescents in higher socioeconomic status families? *Health Place* 2013 Sep;23:79-88 [FREE Full text] [doi: [10.1016/j.healthplace.2013.05.002](https://doi.org/10.1016/j.healthplace.2013.05.002)] [Medline: [23792412](https://pubmed.ncbi.nlm.nih.gov/23792412/)]
53. Larson NI, Story MT, Nelson MC. Neighborhood environments: disparities in access to healthy foods in the US. *Am J Prev Med* 2009 Jan;36(1):74-81. [doi: [10.1016/j.amepre.2008.09.025](https://doi.org/10.1016/j.amepre.2008.09.025)] [Medline: [18977112](https://pubmed.ncbi.nlm.nih.gov/18977112/)]
54. Roth C. Integrating population- and patient-level data for secondary use of electronic health records to study overweight and obesity. *Stud Health Technol Inform* 2013:192.
55. Stanford Natural Language Processing Group. Stanford tokenizer. 2015. URL: <http://nlp.stanford.edu/software/tokenizer.shtml> [accessed 2016-09-28] [WebCite Cache ID 6ks509JgN]
56. Nigam K, Lafferty J, McCallum A. Using maximum entropy for text classification. 1999 Presented at: IJCAI-99 workshop on machine learning for information filtering; 1999; Stockholm, Sweden.
57. Sentiment140 for Academics. URL: <https://sites.google.com/site/twittersentimenthelp/for-researchers> [accessed 2016-08-16] [WebCite Cache ID 6joQzyTSS]
58. Twitter sentiment corpus.: Sanders Analytics; 2011. URL: <http://www.sananalytics.com/lab/twitter-sentiment/> [accessed 2016-08-16] [WebCite Cache ID 6joVh9V4R]
59. Kaggle in Class. 2011. Sentiment Classification URL: <https://inclass.kaggle.com/c/si650winter11> [WebCite Cache ID 6joWjx5fX]
60. National Nutrient Database. Washington, DC: United States Department of Agriculture; 2014 Feb 5. URL: <http://ndb.nal.usda.gov/ndb/search/list?format=&count=&max=25&sort=&fg=&man=&facet=&qlookup=&offset=50> [accessed 2016-09-28]
61. Ainsworth BE. 2011 compendium of physical activities: a second update of codes and MET values. *Med Sci Sport Exer* 2011;43(8):1575-1581.
62. Zhang N. Electronic word of mouth on Twitter about physical activity in the United States: exploratory infodemiology study. *J Med Internet Res* 2013;15(11).
63. Kendall L. Descriptive analysis of physical activity conversations on Twitter. 2011 Presented at: CHI '11 Extended Abstracts on Human Factors in Computing Systems; 2011; Vancouver, Canada p. 1555-1560.
64. Body Measurements.: National Center for Health Statistics, Centers for Disease Control and Prevention; 2012 Sep 2. URL: <http://www.cdc.gov/nchs/fastats/body-measurements.htm> [accessed 2016-09-28] [WebCite Cache ID 6ks62JHO7]
65. Harvard Health Publications. Calories burned in 30 minutes for people of three different weights. 2015. URL: <http://www.health.harvard.edu/newsweek/Calories-burned-in-30-minutes-of-leisure-and-routine-activities.htm> [accessed 2016-09-28] [WebCite Cache ID 6ks6AbOhg]
66. Snow R. Cheap and fast--but is it good? Evaluating non-expert annotations for natural language tasks. 2008 Presented at: Proceedings of the Conference on Empirical Methods in Natural Language Processing; 2008; Stroudsburg, PA p. 254-263.
67. Mitchell L, Frank MR, Harris KD, Dodds PS, Danforth CM. The geography of happiness: connecting twitter sentiment and expression, demographics, and objective characteristics of place. *PLoS One* 2013;8(5):e64417 [FREE Full text] [doi: [10.1371/journal.pone.0064417](https://doi.org/10.1371/journal.pone.0064417)] [Medline: [23734200](https://pubmed.ncbi.nlm.nih.gov/23734200/)]
68. Sentiment140 general information. URL: <http://help.sentiment140.com/> [accessed 2016-09-29] [WebCite Cache ID 6kt10i1Ki]
69. ZIP Code Business Patterns. Washington, DC: US Census Bureau; 2015. URL: <http://www.census.gov/newsroom/press-releases/2015/cb15-tps39.html> [accessed 2016-09-29] [WebCite Cache ID 6kt1J9t9p]
70. Kuczmarski MF, Kuczmarski RJ, Najjar M. Effects of age on validity of self-reported height, weight, and body mass index: findings from the Third National Health and Nutrition Examination Survey, 1988-1994. *J Am Diet Assoc* 2001 Jan;101(1):28-34. [doi: [10.1016/S0002-8223\(01\)00008-6](https://doi.org/10.1016/S0002-8223(01)00008-6)] [Medline: [11209581](https://pubmed.ncbi.nlm.nih.gov/11209581/)]
71. Behavioral Risk Factor Surveillance System Survey Data. Atlanta, GA: Centers for Disease Control and Prevention; 2013. URL: <http://www.cdc.gov/brfss/> [accessed 2016-09-29] [WebCite Cache ID 6kt2xfrVvk]
72. Utah Behavioral Risk Factor Surveillance System Survey Data.: Office of Public Health Assessment, Utah Department of Health; 2014. URL: [http://health.utah.gov/opha/OPHA\\_BRFSS.htm](http://health.utah.gov/opha/OPHA_BRFSS.htm) [accessed 2016-10-06] [WebCite Cache ID 6l4AUrogD]
73. Geographic terms and concepts: census tract.: US Census Bureau; 2012 Jan 06. URL: [https://www.census.gov/geo/reference/gtc/gtc\\_ct.html](https://www.census.gov/geo/reference/gtc/gtc_ct.html) [accessed 2016-10-02] [WebCite Cache ID 6kxgOkozH]



74. 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](https://doi.org/10.2196/jmir.2534)] [Medline: [23989137](https://pubmed.ncbi.nlm.nih.gov/23989137/)]
75. Paul MJ, Derdze M. You are what you tweet: analyzing Twitter for public health. 2011 Jul 05 Presented at: Proceedings of the Fifth International AAI Conference on Weblogs and Social Media; 2011; Barcelona, Spain.
76. De Choudry M, Sharma E, Kiciman E. Characterizing dietary choices, nutrition, and language in food deserts via social media. 2016 Presented at: Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing; 2016; San Francisco, CA p. 1157-1170.
77. McEwen BS. Stress, adaptation, and disease: allostasis and allostatic load. *Ann NY Academy Sci* 1998;840(1):33-44.
78. Seeman TE. Price of adaptation: allostatic load and its health consequences. *Arch Intern Med* 1997;157(19):2259-2268.
79. Duggan M. Social Media Update 2014. Washington, DC: Pew Internet and American Life Project; 2015. URL: [http://www.pewinternet.org/files/2015/01/PI\\_SocialMediaUpdate20144.pdf](http://www.pewinternet.org/files/2015/01/PI_SocialMediaUpdate20144.pdf) [accessed 2016-09-29] [WebCite Cache ID [6kt2HVzYR](https://www.webcitation.org/6kt2HVzYR)]
80. Difference between sample and filter streaming API. 2016 Aug 06. URL: <https://twittercommunity.com/t/difference-between-sample-and-filter-streaming-api/15094> [accessed 2016-09-29] [WebCite Cache ID [6kt2LR5ve](https://www.webcitation.org/6kt2LR5ve)]
81. Burton SH. Right time, right place? Health communication on Twitter: value and accuracy of location information. *J Internet Med Res* 2012;14(6).
82. Morstatter F. Is the Sample Good Enough? Comparing Data from Twitter's Streaming API with Twitter's Firehose. arXiv.5204v1 cs.SI 2013:1306.
83. Burfoot C, Baldwin TA. Automatic satire detection: Are you having a laugh? 2009 Presented at: Proceedings of the Association for Computational Linguistics-IJCNLP; 2009; Singapore.
84. Ptáček T, Habernal I, Hong J. Sarcasm Detection on Czech and English Twitter. In: COLING. 2014 Presented at: 25th International Conference on Computational Linguistics; August 23-29 2014; Dublin, Ireland p. 213-223.
85. Ali MM, Amialchuk A, Heiland FW. Weight-related behavior among adolescents: the role of peer effects. *PLoS One* 2011;6(6):e21179 [FREE Full text] [doi: [10.1371/journal.pone.0021179](https://doi.org/10.1371/journal.pone.0021179)] [Medline: [21731665](https://pubmed.ncbi.nlm.nih.gov/21731665/)]
86. Vartanian LR, Sokol N, Herman CP, Polivy J. Social models provide a norm of appropriate food intake for young women. *PLoS One* 2013;8(11):e79268 [FREE Full text] [doi: [10.1371/journal.pone.0079268](https://doi.org/10.1371/journal.pone.0079268)] [Medline: [24236117](https://pubmed.ncbi.nlm.nih.gov/24236117/)]
87. Cohen DA, Finch BK, Bower A, Sastry N. Collective efficacy and obesity: the potential influence of social factors on health. *Soc Sci Med* 2006 Feb;62(3):769-778. [doi: [10.1016/j.socscimed.2005.06.033](https://doi.org/10.1016/j.socscimed.2005.06.033)] [Medline: [16039767](https://pubmed.ncbi.nlm.nih.gov/16039767/)]
88. Kim D. US state- and county-level social capital in relation to obesity and physical inactivity: A multilevel, multivariable analysis. *Soc Sci Med* 2006;63(4):1045-1059.
89. Berkman L, Syme S. Social networks, host resistance, and mortality: A nine-year follow-up study of Alameda County residents. *Am J Educ* 1979;190(2):186-204.

## Abbreviations

- API:** application programming interface  
**BRFSS:** Behavioral Risk Factor Surveillance System  
**MALLET:** Machine Learning for Language Toolkit  
**Mturk:** Mechanical Turk  
**NAICS:** North American Industry Classification System

*Edited by G Eysenbach; submitted 01.05.16; peer-reviewed by N Zhang, H Zhai, A Jimeno, A MacKinlay, C Seresinhe; comments to author 27.07.16; revised version received 29.08.16; accepted 15.09.16; published 17.10.16.*

*Please cite as:*

Nguyen QC, Li D, Meng HW, Kath S, Nsoesie E, Li F, Wen M  
*Building a National Neighborhood Dataset From Geotagged Twitter Data for Indicators of Happiness, Diet, and Physical Activity*  
*JMIR Public Health Surveill* 2016;2(2):e158  
URL: <http://publichealth.jmir.org/2016/2/e158/>  
doi: [10.2196/publichealth.5869](https://doi.org/10.2196/publichealth.5869)  
PMID: [27751984](https://pubmed.ncbi.nlm.nih.gov/27751984/)

©Quynh C Nguyen, Dapeng Li, Hsien-Wen Meng, Suraj Kath, Elaine Nsoesie, Feifei Li, Ming Wen. Originally published in *JMIR Public Health and Surveillance* (<http://publichealth.jmir.org>), 17.10.2016. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in *JMIR Public Health and Surveillance*, is properly cited. The complete bibliographic information, a link to the original publication on <http://publichealth.jmir.org>, as well as this copyright and license information must be included.



Original Paper

# Evaluating Google, Twitter, and Wikipedia as Tools for Influenza Surveillance Using Bayesian Change Point Analysis: A Comparative Analysis

J Danielle Sharpe<sup>1,2</sup>, MS; Richard S Hopkins<sup>1</sup>, MSPH, MD; Robert L Cook<sup>1</sup>, MPH, MD; Catherine W Striley<sup>1</sup>, MPE, MSW, PhD

<sup>1</sup>College of Public Health and Health Professions, Department of Epidemiology, University of Florida, Gainesville, FL, United States

<sup>2</sup>Rollins School of Public Health, Department of Epidemiology, Emory University, Atlanta, GA, United States

**Corresponding Author:**

J Danielle Sharpe, MS

Rollins School of Public Health

Department of Epidemiology

Emory University

1518 Clifton Road NE

Atlanta, GA, 30322

United States

Phone: 1 912 399 2811

Fax: 1 404 727 8737

Email: [danielle.sharpe@emory.edu](mailto:danielle.sharpe@emory.edu)

## Abstract

**Background:** Traditional influenza surveillance relies on influenza-like illness (ILI) syndrome that is reported by health care providers. It primarily captures individuals who seek medical care and misses those who do not. Recently, Web-based data sources have been studied for application to public health surveillance, as there is a growing number of people who search, post, and tweet about their illnesses before seeking medical care. Existing research has shown some promise of using data from Google, Twitter, and Wikipedia to complement traditional surveillance for ILI. However, past studies have evaluated these Web-based sources individually or dually without comparing all 3 of them, and it would be beneficial to know which of the Web-based sources performs best in order to be considered to complement traditional methods.

**Objective:** The objective of this study is to comparatively analyze Google, Twitter, and Wikipedia by examining which best corresponds with Centers for Disease Control and Prevention (CDC) ILI data. It was hypothesized that Wikipedia will best correspond with CDC ILI data as previous research found it to be least influenced by high media coverage in comparison with Google and Twitter.

**Methods:** Publicly available, deidentified data were collected from the CDC, Google Flu Trends, HealthTweets, and Wikipedia for the 2012-2015 influenza seasons. Bayesian change point analysis was used to detect seasonal changes, or change points, in each of the data sources. Change points in Google, Twitter, and Wikipedia that occurred during the exact week, 1 preceding week, or 1 week after the CDC's change points were compared with the CDC data as the gold standard. All analyses were conducted using the R package "bcp" version 4.0.0 in RStudio version 0.99.484 (RStudio Inc). In addition, sensitivity and positive predictive values (PPV) were calculated for Google, Twitter, and Wikipedia.

**Results:** During the 2012-2015 influenza seasons, a high sensitivity of 92% was found for Google, whereas the PPV for Google was 85%. A low sensitivity of 50% was calculated for Twitter; a low PPV of 43% was found for Twitter also. Wikipedia had the lowest sensitivity of 33% and lowest PPV of 40%.

**Conclusions:** Of the 3 Web-based sources, Google had the best combination of sensitivity and PPV in detecting Bayesian change points in influenza-related data streams. Findings demonstrated that change points in Google, Twitter, and Wikipedia data occasionally aligned well with change points captured in CDC ILI data, yet these sources did not detect all changes in CDC data and should be further studied and developed.

(*JMIR Public Health Surveill* 2016;2(2):e161) doi:[10.2196/publichealth.5901](https://doi.org/10.2196/publichealth.5901)

**KEYWORDS**

Internet; social media; Bayes theorem; public health surveillance; influenza, human

## Introduction

### Background

Although largely vaccine-preventable, influenza places a burden on the US health care system, causing 3000-50,000 deaths annually [1,2]. As one of the many influenza surveillance systems, the Centers for Disease Control and Prevention (CDC) monitors influenza activity by calculating the number of outpatient visits for the syndrome of influenza-like illness (ILI) reported by partnering health care providers to the US Outpatient ILI Surveillance Network (ILINet). The CDC defines ILI as a fever ( $\geq 100^{\circ}\text{F}$  or  $37.8^{\circ}\text{C}$ ) and a cough and sore throat without a known cause other than influenza [3]. This approach to surveillance primarily captures information about people who seek medical care for their influenza symptoms, thus missing those who do not interact with the health care system. In addition, this surveillance method is limited by relatively dated technology and by delays of up to 1 to 2 weeks between the occurrence of the illness event and the dissemination of surveillance information [4].

Syndromic surveillance, which can be defined as the monitoring of disease syndromes in or near real time for early detection of outbreaks, has incorporated the use of novel data sources such as emergency department records and prescription sales to enhance traditional surveillance systems [5-7]. Recently, nontraditional data sources, particularly those that are Web-based, have come into greater application for public health surveillance. This is especially evident as individuals who experience various symptoms may search the Web for health-related information and share their illness experiences using social media platforms before seeking medical care. Using such Web-based data sources such as search queries and social media has been coined digital epidemiology [8-10]. Digital epidemiology can be less expensive, timelier, and can expand detection by increasing the range of health events that can be detected.

### Related Work

As the number of Internet users has increased [11], researchers have identified the use of Google, Twitter, and Wikipedia as novel surveillance approaches to complement traditional methods. Google Flu Trends, which monitors Google users' searches for information related to influenza, has shown correlation with CDC influenza data, while delivering estimates 1 to 2 weeks ahead of CDC reports [8,12]. Although initially successful, the system has not been without its issues in more recent years. Google Flu Trends overestimated influenza activity during the 2012-2013 influenza season and underestimated it during the 2009 H1N1 influenza pandemic [13-16]. One study found that both the original (2008) and revised (2009) algorithms for Google Flu Trends were not reliable on city, regional, and national scales, particularly in instances of varying intensity in influenza seasons and media coverage [16]. Due to issues with its proprietary algorithm, Google Flu Trends was discontinued in August 2015 [17].

Influenza-related posts on Twitter, a social networking platform for disseminating short messages (tweets), have shown high correlation with reported ILI activity in ILINet [18,19]. Studies have found that Twitter data highly correlate with national- and city-level ILI counts [20]. Signorini et al (2011) also demonstrated that tweets could be used to estimate ILI activity at regional and national levels within a reasonable margin of error [21]. Moreover, studies have found that Twitter data perform better than Google data. Nagar et al (2014) conducted a study showing that tweets better reflected city-level ILI incidence in comparison with Google search queries [22]. Aramaki et al discovered that a Twitter-based model outperformed a Google-based model during periods of normal news coverage, although the Twitter model performed less optimally during the periods of excessive media coverage [23]. Moreover, geographic granularity can affect the performance of Twitter data. Broniatowski et al (2015) found that city-level Twitter data performed better than state- and national-level Twitter data, although Google Flu Trends data performed better at each level [24].

Wikipedia page view data have proven valuable for tracking trending topics as well as disease monitoring and forecasting [25,26]. McIver and Brownstein (2014) reported that increases in the quantity of visits to influenza-related Wikipedia articles allowed for the estimation of influenza activity up to 2 weeks before ILINet, outperforming Google Flu Trends estimates during abnormal influenza seasons and periods of high media reporting [27]. One study found that Wikipedia page view data have suitable forecasting value up until the peak of the influenza seasons [26], whereas another study also reported that Wikipedia page view data are suitable for forecasting using a 28-day analysis as well as for nowcasting, or monitoring current disease incidence [25]. However, as a disadvantage, the signal-to-noise ratio of Wikipedia data can be problematic [25] as Wikipedia has become a preferred source for seeking health information whether an individual is ill or not [28,29]. In addition, unlike the granularity flexibility of Google and Twitter data, Wikipedia does not have such capability of evaluating influenza activity at local or regional levels because it only provides counts of page views and no accompanying location or user information in its publicly available data.

### Objective

These early studies on Google, Twitter, and Wikipedia show that, in spite of some drawbacks, mining these Web-based sources may provide valuable epidemic intelligence by identifying indicators of influenza activity at times or in populations that are missed by more traditional surveillance systems. Previous studies have evaluated these 3 Web-based sources individually or dually against a standard, but have not compared all 3 of them with each other and a standard. This comparison is needed to understand if each of these Web-based sources accurately reflect seasonal changes, or change points, that occur in CDC ILI data. It would be beneficial to know which of these Web-based sources performs the best in order

to be considered as a complement to traditional surveillance methods.

Thus, this study aims to conduct a comparative analysis of using Google, Twitter, and Wikipedia for influenza surveillance by examining which Web-based source produces data that are most aligned with CDC ILI data. The specific research question is as follows: For which Web-based source—Google, Twitter, or Wikipedia—do detected change points most closely match change points detected in CDC ILI data for the 2012-2013, 2013-2014, and 2014-2015 influenza seasons? It is hypothesized that Wikipedia data will have the most change points in common with CDC ILI data due to McIver and Brownstein's [27] finding that Wikipedia data can be less influenced by media coverage in comparison with data from Google Flu Trends and Twitter [16,23].

## Methods

### Data Collection

#### Study Period

Data were retrospectively collected for the US-designated 2012-2013 influenza season (September 30, 2012 to May 18, 2013), 2013-2014 influenza season (September 29, 2013 to May 17, 2014), and 2014-2015 influenza season (September 28, 2014 to May 23, 2015) [30-32]. This study period, that is 2012-2015, was chosen due to data constraints. The Twitter data from HealthTweets.org contained tweets dating back to November 2011. As we sought to analyze complete influenza seasons, we could not include the 2011-2012 influenza season, and therefore, any preceding seasons. In addition, we could not include data after the 2014-2015 influenza season because Google ceased making their Google Flu Trends data publicly available in August 2015.

All data were presented as Morbidity and Mortality Weekly Report (MMWR) weeks. MMWR weeks start on Sunday and end on Saturday, ranging from 1 to 52 or 53 weeks [33]. Each of the influenza seasons included in this study begins in MMWR week 40 of a year and ends in week 20 of the following year.

#### CDC Data

Data from the CDC ILINet system were downloaded from FluView Interactive, which provides weekly influenza surveillance information on outpatient illness, hospitalizations, pediatric mortality, virologic surveillance, and geographic activity [34]. ILINet count data are aggregated by MMWR week. The ILINet system aggregates weekly information from participating health care providers on counts of patients seen for ILI by age group, total patients seen by age group, and corresponding year and week [34]. Counts of ILI patient visits to the United States were used for this study. Although most prior studies used weighted ILI rates, we elected to use ILI counts. We decided to use CDC ILI count data to maintain unit comparison because we could not use the Bayesian change point analysis to transpose or model the Web-based count data to a similar scale as the CDC weighted ILI rates.

#### Google Data

Deidentified, national-level count data of influenza-related Google searches made in the United States were downloaded from the Google Flu Trends website [17]. These data are the output of a CDC data-fitted regression model and are based on Google Flu Trends' 2009 model (for the 2012-2013 influenza season), 2013 model (for the 2013-2014 influenza season), and 2014 model (for the 2014-2015 influenza model) [17]. Count data from Google Flu Trends were already aggregated by MMWR week.

#### Twitter Data

For data from Twitter, deidentified, national-level count data of influenza-related tweets in the United States were downloaded from HealthTweets.org, a Johns Hopkins University-based repository of influenza-related tweets dating back to November 2011 [35]. Using the Twitter application programming interface (API), the HealthTweets team collected influenza-related tweets from a keyword stream, which is 1% of public tweets [35]. After collection, Dredze et al [28] categorized the influenza-related tweets using automated annotators based on keywords, keyword combinations, and the classifier developed by Lamb and colleagues [36]. Data from HealthTweets were also already aggregated by MMWR week.

#### Wikipedia Data

Wikipedia has made its article view data available for downloading through Wikimedia Statistics [37]. Wikipedia article view data that are deidentified and aggregated were gathered for views on the "Influenza" article (English version). Count data from the English version of the "Influenza" article served as a proxy for U.S. national-level Wikipedia views. Wikipedia data are presented as the number of article views by the hour, including nonunique views [37]. As Wikipedia article view data on the "Influenza" article are presented by the hour in Wikimedia Statistics, the data were aggregated by MMWR week before analysis.

### Statistical Analysis

#### Bayesian Change Point Analysis

Bayesian change point analysis was the method used for this study. In essence, this technique detects inflections that signal a change within time series data, also known as change points. Bayesian change point analysis has been primarily used to detect when significant changes occur within datasets that have big data properties, such as volume, variety, and velocity [38]. For instance, Bayesian change point analysis has been used to estimate when changes occurred in interest rate data [39], chromosomal microarray data [39], and cancer-related gene expression data [40]. This method was used to detect changes in emergency department attendance and hospital admissions after a health system transformation in a post-earthquake area [41]. Bayesian change point analysis has also been used to detect changes in the dynamics of an aquatic ecosystem such as the introduction of a nonnative species [42]. Besides our study, this Bayesian technique has been used only once for influenza surveillance using ILI visits to emergency departments [43], which is unlike our analysis in that we used Web-based data.

**Figure 1.** Simplified equation by Barry and Hartigan.

$$\frac{p_i}{1 - p_i} = \frac{P(U_i = 1|X, U_j, j \neq i)}{P(U_i = 0|X, U_j, j \neq i)} = \frac{\left[ \int_0^\gamma p^b (1-p)^{n-b-1} dp \right] \left[ \int_0^\lambda \frac{\omega^{b/2}}{(W_1 + B_1 \omega)^{(n-1)/2}} d\omega \right]}{\left[ \int_0^\gamma p^{b-1} (1-p)^{n-b} dp \right] \left[ \int_0^\lambda \frac{\omega^{(b-1)/2}}{(W_0 + B_0 \omega)^{(n-1)/2}} d\omega \right]}$$

where  $P$  is the probability of a change point;  $i$  is a position or single interval in time series data;  $j$  is the ending position;  $U_i=1$  is a change point at position  $i + 1$ ;  $X$  is the data; parameters  $\gamma$  and  $\lambda$  are values between 0 and 1;  $b$  is the number of blocks found if  $U_i=0$ , conditional on  $U_j$ , for  $i \neq j$ ; and  $W_0, B_0, W_1$ , and  $B_1$  are the within and between block sums of squares found when  $U_i=0$  and  $U_i=1$ , respectively

Bayesian change point analysis formed the method of choice as it is one of the proven methods that can detect subtle changes in time series data more effectively than traditional aberration detection methods [43]. Kass-Hout et al (2012) found that Bayesian change point analysis was not as sensitive as 2 other change point analysis methods—the cumulative sum technique and structural change model [43]. However, Bayesian change point analysis has been best applied to microarray data [39,40], which have big data properties similar to Web-based data.

All Bayesian change point analyses were conducted using the R package “bcp” version 4.0.0 [39,40,44] in RStudio version 0.99.484 [45]. The “bcp” package implements a complex Markov Chain Monte Carlo (MCMC) approximation [39,40,44] of the Bayesian change point method described by Barry and Hartigan [46]. As the default for the “bcp” package, after 500 MCMC iterations, the probability of a change point at any given interval (ie, MMWR week) in time series data is computed from the number of times in the MCMC iterations that the condition of having a change point at that interval was met [39,40,44].

In each step of the Markov chain, the transition probability,  $p$ , for the conditional probability of a change point is found from the simplified equation by Barry and Hartigan [39,40,46], which is provided in Figure 1. After each MCMC iteration, the posterior means and probabilities are updated until the end of the time series. It is recommended that readers refer to Erdman and Emerson (2007), Erdman and Emerson (2008), and Barry and Hartigan (1993) for further mathematical explanation of this Bayesian method [39,40,46].

### Change Points

We considered significant change points to be where the Bayesian method indicated the probability of a change occurring as  $\geq 50\%$ . Change points detected in the CDC ILI data were the gold standard with which change points found in the Web-based sources were compared. Change points of the Web-based sources that occurred during the exact week, 1 preceding week, or 1 week after the CDC change points were considered matching or true change points. This was done to account for

any reporting lags that can be common with surveillance data. The number of change points for each data source was compared, and sensitivity and PPV for the detection of change points were calculated for each of the Web-based sources.

### Sensitivity and Positive Predictive Values

Sensitivity and PPV were computed for each Web-based source using the change points detected for that Web-based source that matched change points detected for the CDC ILINet system (true positives), change points detected for the ILINet system but not for the Web-based source (false negatives), and change points detected for the Web-based source but not for the ILINet system (false positives). Sensitivity was calculated by dividing the true positives for each Web-based source by the total of true positives and false negatives, which would be the total number of CDC change points [47]. PPV were calculated by dividing the true positives for each Web-based source by the total of true positives and false positives, which would be the total change points for that particular Web-based source [47].

## Results

### Sample Characteristics

A summary of the count data that were collected and analyzed for the CDC, Google, Twitter, and Wikipedia for the 2012-2015 influenza seasons is provided (Table 1). There was year-to-year variability in the average weekly counts of events included for each of the data sources. For most of the data sources, the 2012-2013 influenza season had the highest average number of weekly counts. The 2013-2014 influenza season had the lowest average number for the CDC and Google Flu Trends, whereas the 2014-2015 influenza season had the lowest average number for Twitter and Wikipedia. Note that the 2014-2015 influenza season consisted of 34 total MMWR weeks because whereas most epidemiologic years are comprised of 52 MMWR weeks, the 2014-2015 epidemiologic year had 53 weeks due to a preceding calendar leap year. Table 1 further summarizes the data information.



**Table 1.** Summary of weekly Influenza-like Illness count data for the Centers for Disease Control and Prevention, Google, Twitter, and Wikipedia, 2012-2015 influenza seasons.

Influenza season	CDC <sup>a</sup> ILINet <sup>b</sup>	Google	Twitter	Wikipedia
<b>2012-2013</b>				
MMWR <sup>c</sup> Weeks (counts/week)	33	33	33	33
Mean	19,049	4121	8096	47,541
Min	7317	1286	2558	29,865
Max	39,896	10,555	22,935	114,919
<b>2013-2014</b>				
MMWR Weeks (counts/week)	33	33	33	33
Mean	16,574	2274	5826	25,039
Min	9033	1339	1196	17,885
Max	28,654	5008	10,506	36,935
<b>2014-2015</b>				
MMWR Weeks (counts/week)	34	34	34	34
Mean	19,940	2549	2900	21,918
Min	9289	1144	451	12,958
Max	40,664	6911	8709	35,232

<sup>a</sup>CDC: Centers for Disease Control and Prevention.

<sup>b</sup>ILINet: United States Outpatient Influenza-like Illness Surveillance Network.

<sup>c</sup>MMWR: Morbidity and Mortality Weekly Report.

### Comparison of Change Points Detected in the 2012-2015 Influenza Seasons

A summary of all change points found in each data source is provided (see [Figures 2-4](#)), and a comparison of change points is shown in [Table 2](#). For the 2012-2013 influenza season, Google had 3 total change points in common with the CDC ILINet system, which were MMWR weeks 51, 4, and 5. Twitter had 2 change points in common with the CDC's change points, which were MMWR weeks 47 and 4. Wikipedia had only 1 change point that matched the CDC ILINet system, which was MMWR week 5.

In the 2013-2014 influenza season, Google had a total of 4 change points (MMWR weeks 48, 50, 51, and 5) that coincided with change points detected in the CDC data. Twitter had 3 change points (MMWR weeks 48, 51, and 7) that matched change points in the CDC ILINet system's data. Wikipedia had 2 change points in common with CDC ILI data, which were MMWR weeks 51 and 6.

For the 2014-2015 influenza season, 4 change points (MMWR weeks 48, 50, 51, and 53) were detected in the Google data that concurred with change points identified in the CDC ILINet system. Both Twitter and Wikipedia had only 1 change point that coincided with the change points found in the CDC ILI data, which were MMWR weeks 50 and 53, respectively.

### Comparison of Sensitivity and Positive Predictive Value Detected Among Web-Based Sources

Next, we computed the sensitivity and PPV for each of the Web-based sources using the CDC ILI data as the gold standard. As shown in [Table 3](#), results varied widely across the Web-based sources. A high sensitivity of 92% was found for Google, while the PPV for Google was 85%. A low sensitivity of 50% was calculated for Twitter; a low PPV of 43% was found for Twitter also. Wikipedia had the lowest sensitivity of 33% and lowest PPV of 40%. A table comparing sensitivity and PPV by specific influenza season is also provided (see [Multimedia Appendix 1](#)).



**Table 2.** Comparison of change points detected using Bayesian change point analysis, 2012-2015 influenza seasons<sup>a</sup>.

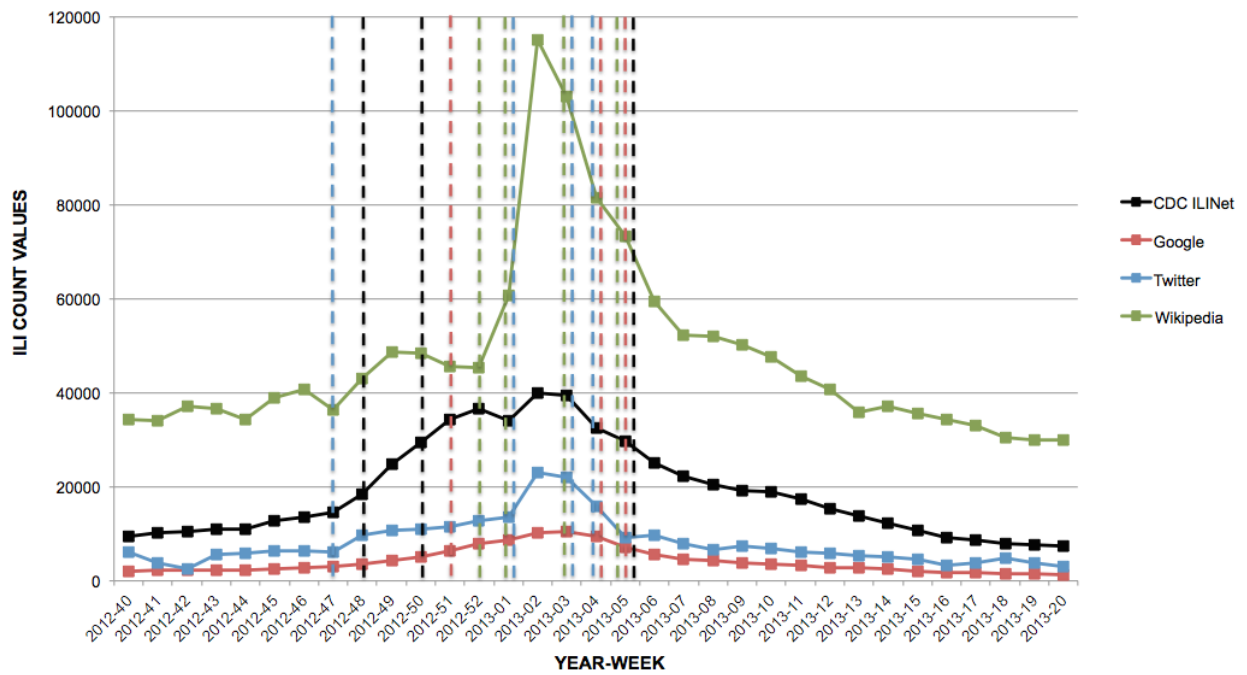
Influenza season	CDC <sup>b</sup> ILINet <sup>c</sup> counts (reference)	Google counts	Twitter counts	Wikipedia counts
<b>2012-2013</b>			Week 47 <sup>a</sup>	
	Week 48			
	Week 50			
		Week 51 <sup>a</sup>		
				Week 52
			Week 1	Week 1
			Week 3	Week 3
		Week 4 <sup>a</sup>	Week 4 <sup>a</sup>	
	Week 5	Week 5 <sup>a</sup>		Week 5 <sup>a</sup>
	<b>2013-2014</b>	Week 48	Week 48 <sup>a</sup>	Week 48 <sup>a</sup>
Week 50		Week 50 <sup>a</sup>		
		Week 51 <sup>a</sup>	Week 51 <sup>a</sup>	Week 51 <sup>a</sup>
				Week 1
		Week 3		
			Week 4	
		Week 5 <sup>a</sup>		
Week 6				Week 6 <sup>a</sup>
			Week 7 <sup>a</sup>	
Week 15				
<b>2014-2015</b>			Week 17	
			Week 43	
				Week 44
	Week 48	Week 48 <sup>a</sup>		
	Week 49			
	Week 50	Week 50 <sup>a</sup>	Week 50 <sup>a</sup>	
		Week 51 <sup>a</sup>		
	Week 53	Week 53 <sup>a</sup>		Week 53 <sup>a</sup>
			Week 2	
			Week 3	Week 3
Week 6		Week 4		
		Week 12		

<sup>a</sup>MMWR week indicates a corresponding change point to the CDC change points (reference).

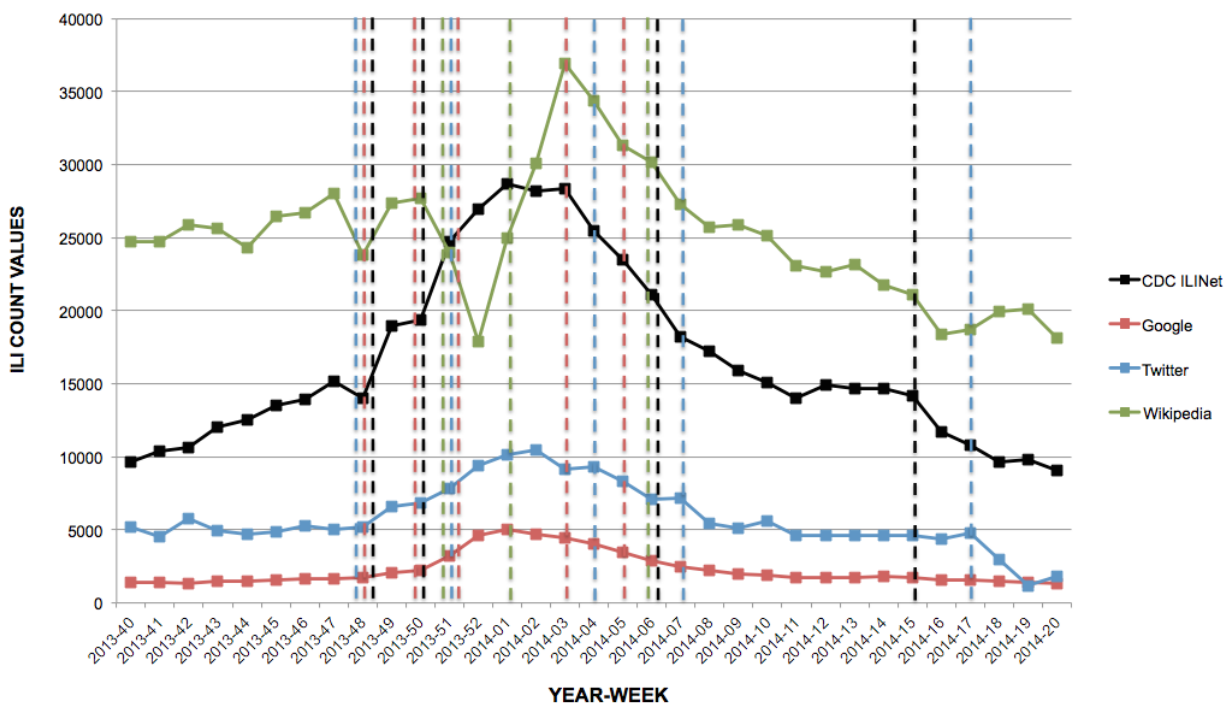
<sup>b</sup>CDC: Centers for Disease Control and Prevention.

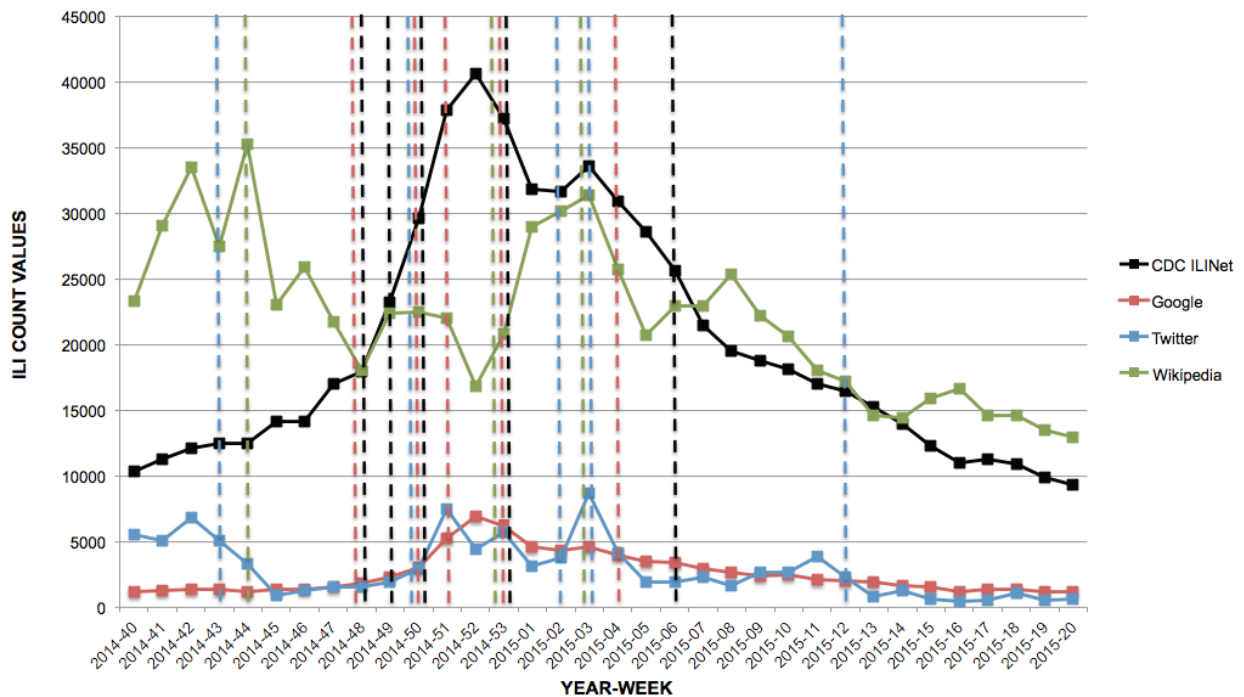
<sup>c</sup>ILINet: United States Outpatient Influenza-like Illness Surveillance Network.

**Figure 2.** Change points (dotted lines) detected by Bayesian change point analysis, 2012-2013 influenza season.



**Figure 3.** Change points (dotted lines) detected by Bayesian change point analysis, 2013-2014 influenza season.



**Figure 4.** Change points (dotted lines) detected by Bayesian change point analysis, 2014-2015 influenza season.**Table 3.** Comparison of sensitivity and positive predictive value among Web-based sources, 2012-2015 influenza seasons.

Web-based source	Sensitivity (%)	Positive predictive value (%)
Google	92	85
Twitter	50	43
Wikipedia	33	40

## Discussion

### Principal Findings

Google had a total of 11 true change points (3 in the 2012-2013 influenza season, 4 in the 2013-2014 influenza season, and 4 in the 2014-2015 influenza season) that coincided with the CDC ILINet's change points. As Google had the most change points that coincided with change points detected in the CDC ILI data, our hypothesis that Wikipedia would have the most change points was not supported. Sensitivity and PPV for event detection are important for evaluating the quality of surveillance systems [47]. Google had a moderate positive predictive value and was highly sensitive, whereas Twitter and Wikipedia both had low sensitivity rates and PPVs. This finding that Google had the best correspondence is not consistent with that of the previous studies that have found Twitter and Wikipedia to perform better [22,23,27,48].

Google, Twitter, and Wikipedia all had some change points that aligned well with CDC ILI data; however, they did not identify all change points that were identified in the CDC data, which would be important for understanding when seasonal changes occur during an influenza season. As no Web-based source identified all detected changes in the CDC data, this could indicate that the Web-based data, itself, may be limited in capturing all changes of CDC ILI data, which is quite plausible

as not every individual who experiences ILI symptoms resorts to searching or sharing health information online. On the contrary, this could indicate that the Bayesian change point analysis as a technique is not adequately sensitive for the use on Web-based data. These 3 Web-based sources need to be further studied and compared using more standard statistical methods before being incorporated as surveillance data to complement a traditional system.

### Limitations

There are limitations of this study that should be noted. First, Bayesian change point analysis assumes time series data are distributed normally, which can be problematic as public health surveillance data can be variable and can have a nonnormal distribution [43]. However, we were unable to test this assumption on the "bcp" package in RStudio, and this is a limitation because the "bcp" package could have incorrectly identified or missed change points, especially if there were any outliers in the data to skew the Bayesian analysis. Another major limitation to using Bayesian change point analysis is that it cannot be used as a technique to monitor real-time data [49]. Bayesian change point analysis is best used to evaluate changes in historical time series data after all data have been collected. For this study, the Bayesian method was used to retrospectively evaluate data collected from the CDC, Google, Twitter, and Wikipedia after each influenza season occurred; therefore, the

results cannot be directly applied for prospective use or real-time influenza surveillance.

A possible solution to conducting real-time influenza surveillance using Web-based data may lie in using a normal distribution algorithm. Normal distribution methods that are based on historical limits and cumulative sums have been traditionally used for influenza surveillance by the CDC [50]. Moreover, Pervaiz et al (2012) demonstrated that real-time influenza surveillance using Web-based data could be done more effectively using negative binomial- and Poisson-based models as opposed to normal distribution models due to the noisy nature of Web-based data and fluctuating numbers of Internet users and their activity levels [50].

Second, for the analysis of Wikipedia views, only the “Influenza” article was used for analysis, excluding other articles on influenza medications and influenza strains. McIver and Brownstein described the effectiveness of combining multiple influenza-related Wikipedia articles for surveillance purposes [27], but those were not included in this study. We assumed all views of the English-language Wikipedia “Influenza” article were by US users; however, some may have come from users in other English-speaking countries where the influenza season is very different, such as Australia.

Third, some of our data sources may have limits. We used CDC ILI count data for the analysis, which is not standard. Most prior studies have used weighted ILI rates instead of ILI counts because the weighted rates account for population variations in the United States. Using ILI counts may have sampling biases, but we justify the use of counts because we wanted to maintain data uniformity as none of the Web-based count data accounted for or could be normalized by population and regional variations in the United States. In addition, the Bayesian change point analysis did not allow us to transpose Web-based count data on the same scale as weighted ILI rates, thus ILI counts were the best option, considering the method used. Furthermore, the Google Flu Trends data used in this study were the output of a regression model that was fitted to CDC ILI data, leading to the Google data being a closer comparison with CDC ILI data. Although the Google Flu Trends data were fitted to match CDC data, it is important to note that these were readily available to the public as well as practitioners, justifying their use.

Fourth, data duplication could be an issue with each data source used in this study. Internet users can use a single website for multiple information searches and shares, and a single Internet user can use multiple websites for the same information search or share [51]. For example, a user can view the Wikipedia “Influenza” article multiple times and each view would be considered as a separate count [37]. Neither Google Flu Trends nor HealthTweets can distinguish or remove multiple searches and tweets by a single user [12,35]. In addition, there is no way,

in publicly available data, to distinguish when a single user searches both Wikipedia and Google for the same information. The CDC ILINet system does not differentiate when a single patient makes repeated outpatient visits to the same participating health care provider or when a single patient makes outpatient visits to multiple health care providers for the same illness. This issue of data duplication should be further investigated in future studies.

Finally, Internet users are, on average, younger than the general U.S. population [52]. Although this difference may be viewed as a limitation to using Web-based data for influenza surveillance, younger age groups (0-4 years, 5-24 years, and 25-49 years) account for a majority of the outpatient ILI counts that are reported to the CDC ILINet system [34].

### Future Research

There is more substantive information in the content of Web-based sources that is not accounted for in count data of Web-based sources. Recent research has already begun to conduct content analyses of Web-based sources such as chat forums, Facebook, and Twitter in order to understand the health experiences and needs addressed by Internet users. Content analyses have proven valuable for both communicable and noncommunicable diseases because Internet users share and search about various health experiences ranging from mental health [53,54] to substance use [55,56] to the health needs of sexual minorities [57]. In addition, public health surveillance can be strengthened by combining various data sources, whether Web-based or traditional. Santillana et al (2015) found that when data from Google, Twitter, hospital records, and a participatory surveillance system were combined, influenza activity was predicted more accurately than and up to 4 weeks before the CDC [58]. More research should be carried out in this area to identify the best combination of traditional and novel data sources for influenza surveillance.

### Conclusions

To our knowledge, this is the first comparison to evaluate Google, Twitter, and Wikipedia as possible data sources for influenza surveillance against a common gold standard (the CDC ILINet system). Of the 3 Web-based sources, Google had the best combination of sensitivity and PPV in detecting Bayesian change points in influenza-related data streams. This finding is not consistent with existing research that has compared Google and Twitter data or Google and Wikipedia data, which could be attributed to the analysis of different influenza seasons, the novel use of the Bayesian method in this study, or the fact that Google Flu Trends data were fitted to CDC data. Further research should assess the substantive health content contained within these 3 Web-based sources, the surveillance value of combining these sources, and the ability of these sources to detect influenza activity using other statistical methods.

### Acknowledgments

The authors thank Dr. Heather Rubino (Florida Department of Health), Sophie Smith (Centers for Disease Control and Prevention), and Matthew Biggerstaff (Centers for Disease Control and Prevention) for their insight and expertise with understanding current influenza surveillance systems. We also acknowledge Dr. John Emerson (Yale University), Dr. Chandra Erdman (Management

Leadership for Tomorrow), and Dr. Xiaofei Wang (Amherst College) for their assistance with understanding the R package “bcp” version 4.0.0. Finally, we extend special appreciation to Google Inc, Dr. Mark Dredze and his team at HealthTweets.org, and Wikimedia Statistics for making their data publicly accessible.

### Authors' Contributions

JDS conceptualized and designed the study. RLC, RSH, and CWS provided feedback for the study design. JDS acquired data and performed the statistical analysis. JDS, RLC, RSH, and CWS contributed to the interpretation of findings. JDS drafted and revised the manuscript according to feedback provided by RLC, RSH, and CWS. All authors approved the final version of the manuscript.

### Conflicts of Interest

None declared.

### Multimedia Appendix 1

Comparison of sensitivity and positive predictive value (PPV) among web-based sources by specific influenza season.

[\[PDF File \(Adobe PDF File\), 105KB - publichealth\\_v2i2e161\\_app1.pdf\]](#)

### References

1. Chunara R, Goldstein E, Patterson-Lomba O, Brownstein JS. Estimating influenza attack rates in the United States using a participatory cohort. *Sci Rep* 2015;5:9540 [FREE Full text] [doi: [10.1038/srep09540](https://doi.org/10.1038/srep09540)] [Medline: [25835538](https://pubmed.ncbi.nlm.nih.gov/25835538/)]
2. Centers for Disease Control and Prevention. Estimates of deaths associated with seasonal influenza --- United States, 1976-2007. *MMWR Morb Mortal Wkly Rep* 2010;59(33):1057-1062. [Medline: [20798667](https://pubmed.ncbi.nlm.nih.gov/20798667/)]
3. Centers for Disease Control and Prevention. 2015. Overview of influenza surveillance in the United States URL: <http://www.cdc.gov/flu/weekly/overview.htm> [accessed 2016-03-26] [WebCite Cache ID [6gHxdKs79](#)]
4. Fairchild G, De Silva L, Del Valle S, Segre A. Eliciting disease data from Wikipedia articles. 2015 Presented at: Ninth International AAAI Conference on Web and Social Media; 2015; Oxford, England p. 26-33.
5. Buehler JW, Hopkins RS, Overhage JM, Sosin DM, Tong V. Framework for evaluating public health surveillance systems for early detection of outbreaks: recommendations from the CDC Working Group. *MMWR Recomm Rep* 2004 May 7;53(RR-5):1-11 [FREE Full text] [Medline: [15129191](https://pubmed.ncbi.nlm.nih.gov/15129191/)]
6. Buehler JW, Sonricker A, Paladini M, Soper P, Mostashari F. Syndromic surveillance practice in the United States: findings from a survey of state, territorial, and selected local health departments. *Adv Dis Surveill* 2008;6(3):1-20.
7. Henning KJ. What is syndromic surveillance? *MMWR Suppl* 2004 Sep 24;53:5-11 [FREE Full text] [Medline: [15714620](https://pubmed.ncbi.nlm.nih.gov/15714620/)]
8. Bernardo TM, Rajic A, Young I, Robiadek K, Pham MT, Funk JA. Scoping review on search queries and social media for disease surveillance: a chronology of innovation. *J Med Internet Res* 2013;15(7):e147 [FREE Full text] [doi: [10.2196/jmir.2740](https://doi.org/10.2196/jmir.2740)] [Medline: [23896182](https://pubmed.ncbi.nlm.nih.gov/23896182/)]
9. 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](https://doi.org/10.2196/jmir.1157)] [Medline: [19329408](https://pubmed.ncbi.nlm.nih.gov/19329408/)]
10. 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](https://doi.org/10.1371/journal.pcbi.1002616)] [Medline: [22844241](https://pubmed.ncbi.nlm.nih.gov/22844241/)]
11. International Telecommunication Union. ICT facts and figures: the world in 2015 URL: <http://www.itu.int/en/ITU-D/Statistics/Documents/facts/ICTFactsFigures2015.pdf> [accessed 2016-03-26] [WebCite Cache ID [6gHxY2tQX](#)]
12. Ginsberg J, Mohebbi MH, Patel RS, Brammer L, Smolinski MS, Brilliant L. Detecting influenza epidemics using search engine query data. *Nature* 2009 Feb 19;457(7232):1012-1014. [doi: [10.1038/nature07634](https://doi.org/10.1038/nature07634)] [Medline: [19020500](https://pubmed.ncbi.nlm.nih.gov/19020500/)]
13. Cook S, Conrad C, Fowlkes AL, Mohebbi MH. Assessing Google flu trends performance in the United States during the 2009 influenza virus A (H1N1) pandemic. *PLoS One* 2011;6(8):e23610 [FREE Full text] [doi: [10.1371/journal.pone.0023610](https://doi.org/10.1371/journal.pone.0023610)] [Medline: [21886802](https://pubmed.ncbi.nlm.nih.gov/21886802/)]
14. Butler D. When Google got flu wrong. *Nature* 2013 Feb 14;494(7436):155-156. [doi: [10.1038/494155a](https://doi.org/10.1038/494155a)] [Medline: [23407515](https://pubmed.ncbi.nlm.nih.gov/23407515/)]
15. 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](https://doi.org/10.1126/science.1248506)] [Medline: [24626916](https://pubmed.ncbi.nlm.nih.gov/24626916/)]
16. Olson DR, Konty KJ, Paladini M, Viboud C, Simonsen L. Reassessing Google Flu Trends data for detection of seasonal and pandemic influenza: a comparative epidemiological study at three geographic scales. *PLoS Comput Biol* 2013;9(10):e1003256 [FREE Full text] [doi: [10.1371/journal.pcbi.1003256](https://doi.org/10.1371/journal.pcbi.1003256)] [Medline: [24146603](https://pubmed.ncbi.nlm.nih.gov/24146603/)]
17. Google. Google Flu Trends URL: <http://www.google.org/flutrends/about/> [accessed 2016-03-26] [WebCite Cache ID [6gHx8E3SD](#)]
18. Broniatowski DA, Paul MJ, Dredze M. Twitter: big data opportunities. *Science* 2014 Jul 11;345(6193):148. [doi: [10.1126/science.345.6193.148-a](https://doi.org/10.1126/science.345.6193.148-a)] [Medline: [25013052](https://pubmed.ncbi.nlm.nih.gov/25013052/)]



19. Aslam AA, Tsou M, Spitzberg BH, An L, Gawron JM, Gupta DK, et al. The reliability of tweets as a supplementary method of seasonal influenza surveillance. *J Med Internet Res* 2014;16(11):e250 [FREE Full text] [doi: [10.2196/jmir.3532](https://doi.org/10.2196/jmir.3532)] [Medline: [25406040](https://pubmed.ncbi.nlm.nih.gov/25406040/)]
20. Broniatowski DA, Paul MJ, Dredze M. National and local influenza surveillance through Twitter: an analysis of the 2012-2013 influenza epidemic. *PLoS One* 2013;8(12):e83672 [FREE Full text] [doi: [10.1371/journal.pone.0083672](https://doi.org/10.1371/journal.pone.0083672)] [Medline: [24349542](https://pubmed.ncbi.nlm.nih.gov/24349542/)]
21. Signorini A, Segre AM, Polgreen PM. The use of Twitter to track levels of disease activity and public concern in the U.S. during the influenza A H1N1 pandemic. *PLoS One* 2011;6(5):e19467 [FREE Full text] [doi: [10.1371/journal.pone.0019467](https://doi.org/10.1371/journal.pone.0019467)] [Medline: [21573238](https://pubmed.ncbi.nlm.nih.gov/21573238/)]
22. Nagar R, Yuan Q, Freifeld CC, Santillana M, Nojima A, Chunara R, et al. A case study of the New York City 2012-2013 influenza season with daily geocoded Twitter data from temporal and spatiotemporal perspectives. *J Med Internet Res* 2014;16(10):e236 [FREE Full text] [doi: [10.2196/jmir.3416](https://doi.org/10.2196/jmir.3416)] [Medline: [25331122](https://pubmed.ncbi.nlm.nih.gov/25331122/)]
23. Aramaki E, Maskawa S, Morita M. Twitter catches the flu: detecting influenza epidemics using Twitter. 2011 Presented at: Conference on Empirical Natural Language Processing Conference (EMNLP); 2011; Edinburgh, Scotland p. 1568-1576.
24. Broniatowski DA, Dredze M, Paul MJ, Dugas A. Using Social Media to Perform Local Influenza Surveillance in an Inner-City Hospital: A Retrospective Observational Study. *JMIR Public Health Surveill* 2015;1(1):e5 [FREE Full text] [doi: [10.2196/publichealth.4472](https://doi.org/10.2196/publichealth.4472)] [Medline: [27014744](https://pubmed.ncbi.nlm.nih.gov/27014744/)]
25. Generous N, Fairchild G, Deshpande A, Del Valle SY, Priedhorsky R. Global disease monitoring and forecasting with Wikipedia. *PLoS Comput Biol* 2014 Nov;10(11):e1003892 [FREE Full text] [doi: [10.1371/journal.pcbi.1003892](https://doi.org/10.1371/journal.pcbi.1003892)] [Medline: [25392913](https://pubmed.ncbi.nlm.nih.gov/25392913/)]
26. Hickmann KS, Fairchild G, Priedhorsky R, Generous N, Hyman JM, Deshpande A, et al. Forecasting the 2013-2014 influenza season using Wikipedia. *PLoS Comput Biol* 2015 May;11(5):e1004239 [FREE Full text] [doi: [10.1371/journal.pcbi.1004239](https://doi.org/10.1371/journal.pcbi.1004239)] [Medline: [25974758](https://pubmed.ncbi.nlm.nih.gov/25974758/)]
27. 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](https://doi.org/10.1371/journal.pcbi.1003581)] [Medline: [24743682](https://pubmed.ncbi.nlm.nih.gov/24743682/)]
28. Laurent MR, Vickers TJ. Seeking health information online: does Wikipedia matter? *J Am Med Inform Assoc* 2009;16(4):471-479 [FREE Full text] [doi: [10.1197/jamia.M3059](https://doi.org/10.1197/jamia.M3059)] [Medline: [19390105](https://pubmed.ncbi.nlm.nih.gov/19390105/)]
29. Tausczik Y, Faasse K, Pennebaker JW, Petrie KJ. Public anxiety and information seeking following the H1N1 outbreak: blogs, newspaper articles, and Wikipedia visits. *Health Commun* 2012;27(2):179-185. [doi: [10.1080/10410236.2011.571759](https://doi.org/10.1080/10410236.2011.571759)] [Medline: [21827326](https://pubmed.ncbi.nlm.nih.gov/21827326/)]
30. Centers for Disease Control and Prevention. Influenza activity - United States, 2012-2013 season and composition of the 2013-14 influenza vaccine. *MMWR Morb Mortal Wkly Rep* 2013;62(23):473-479. [Medline: [23760189](https://pubmed.ncbi.nlm.nih.gov/23760189/)]
31. Epperson S, Blanton L, Kniss K, Mustaquim D, Steffens C, Wallis T, Influenza Division, National Center for Immunization and Respiratory Diseases, CDC. Influenza activity - United States, 2013-14 season and composition of the 2014-15 influenza vaccines. *MMWR Morb Mortal Wkly Rep* 2014 Jun 6;63(22):483-490 [FREE Full text] [Medline: [24898165](https://pubmed.ncbi.nlm.nih.gov/24898165/)]
32. Appiah GD, Blanton L, D'Mello T, Kniss K, Smith S, Mustaquim D, Centers for Disease Control and Prevention (CDC). Influenza activity - United States, 2014-15 season and composition of the 2015-16 influenza vaccine. *MMWR Morb Mortal Wkly Rep* 2015 Jun 5;64(21):583-590 [FREE Full text] [Medline: [26042650](https://pubmed.ncbi.nlm.nih.gov/26042650/)]
33. Centers for Disease Control and Prevention. MMWR weeks URL: [https://wwwn.cdc.gov/443/nndss/document/MMWR\\_Week\\_overview.pdf](https://wwwn.cdc.gov/443/nndss/document/MMWR_Week_overview.pdf) [accessed 2016-03-26] [WebCite Cache ID 6gHxPOqx0]
34. Centers for Disease Control and Prevention. FluView Interactive URL: <http://gis.cdc.gov/grasp/fluview/fluportaldashboard.html> [accessed 2016-03-26] [WebCite Cache ID 6gHxKR4m4]
35. Dredze M, Cheng R, Paul M, Broniatowski D. HealthTweets.org: a platform for public health surveillance using Twitter. 2014 Presented at: Twenty-Eighth AAAI Conference on Artificial Intelligence; 2014; Quebec City, Quebec, Canada.
36. Lamb A, Paul M, Dredze M. Separating fact from fear: tracking flu infections on Twitter. In: Proceedings of the 2013 Conference of the North American Chapter of the Association for Computational Linguistics (NAACL): Human Language Technologies. 2013 Presented at: Conference of the North American Chapter of the Association for Computational Linguistics (NAACL): Human Language Technologies; 2013; Atlanta, GA p. 789-795.
37. Wikimedia. Page view statistics for Wikimedia projects URL: <http://dumps.wikimedia.org/other/pagecounts-raw/> [accessed 2016-03-26] [WebCite Cache ID 6gHwztskJ]
38. De Mauro A, Greco M, Grimaldi M. What is big data? A consensual definition and a review of key research topics. In: Proceedings of AIP Conference.: AIP Publishing LLC; 2015 Presented at: 4th International Conference on Integrated Information; September 5-8, 2014; Madrid, Spain p. 97-104. [doi: [10.1063/1.4907823](https://doi.org/10.1063/1.4907823)]
39. Erdman C, Emerson JW. bcp: an R package for performing a Bayesian analysis of change point problems. *J. Stat. Soft* 2007;23(3). [doi: [10.18637/jss.v023.i03](https://doi.org/10.18637/jss.v023.i03)]
40. Erdman C, Emerson JW. A fast Bayesian change point analysis for the segmentation of microarray data. *Bioinformatics* 2008 Oct 1;24(19):2143-2148 [FREE Full text] [doi: [10.1093/bioinformatics/btn404](https://doi.org/10.1093/bioinformatics/btn404)] [Medline: [18667443](https://pubmed.ncbi.nlm.nih.gov/18667443/)]

41. Schluter PJ, Hamilton GJ, Deely JM, Ardagh MW. Impact of integrated health system changes, accelerated due to an earthquake, on emergency department attendances and acute admissions: a Bayesian change-point analysis. *BMJ Open* 2016;6(5):e010709 [FREE Full text] [doi: [10.1136/bmjopen-2015-010709](https://doi.org/10.1136/bmjopen-2015-010709)] [Medline: [27169741](https://pubmed.ncbi.nlm.nih.gov/27169741/)]
42. Ellis BK, Stanford JA, Goodman D, Stafford CP, Gustafson DL, Beauchamp DA, et al. Long-term effects of a trophic cascade in a large lake ecosystem. *Proc Natl Acad Sci U S A* 2011 Jan 18;108(3):1070-1075 [FREE Full text] [doi: [10.1073/pnas.1013006108](https://doi.org/10.1073/pnas.1013006108)] [Medline: [21199944](https://pubmed.ncbi.nlm.nih.gov/21199944/)]
43. Kass-Hout TA, Xu Z, McMurray P, Park S, Buckeridge DL, Brownstein JS, et al. Application of change point analysis to daily influenza-like illness emergency department visits. *J Am Med Inform Assoc* 2012;19(6):1075-1081 [FREE Full text] [doi: [10.1136/amiajnl-2011-000793](https://doi.org/10.1136/amiajnl-2011-000793)] [Medline: [22759619](https://pubmed.ncbi.nlm.nih.gov/22759619/)]
44. Wang X, Emerson J. arxiv. 2015. Bayesian change point analysis of linear models on general graphs URL: <https://arxiv.org/pdf/1509.00817> [accessed 2016-10-17] [WebCite Cache ID 6JJxaahQr]
45. Rstudio. Boston, MA: RStudio, Inc; 2015. RStudio: integrated development for R URL: <https://www.rstudio.com/> [accessed 2016-03-26] [WebCite Cache ID 6gHzyHxHu]
46. Barry D, Hartigan J. A Bayesian analysis for change point problems. *JASA* 1993;88(421):309-319.
47. German RR, Lee LM, Horan JM, Milstein RL, Pertowski CA, Waller MN, Guidelines Working Group Centers for Disease Control and Prevention (CDC). Updated guidelines for evaluating public health surveillance systems: recommendations from the Guidelines Working Group. *MMWR Recomm Rep* 2001 Jul 27;50(RR-13):1-35; quiz CE1-7. [Medline: [18634202](https://pubmed.ncbi.nlm.nih.gov/18634202/)]
48. Paul MJ, Dredze M, Broniatowski D. Twitter improves influenza forecasting. *PLoS Curr* 2014;6 [FREE Full text] [doi: [10.1371/currents.outbreaks.90b9ed0f59bae4ccaa683a39865d9117](https://doi.org/10.1371/currents.outbreaks.90b9ed0f59bae4ccaa683a39865d9117)] [Medline: [25642377](https://pubmed.ncbi.nlm.nih.gov/25642377/)]
49. Kass-Hout T. Google. 2013. Change point analysis: Introduction URL: <https://sites.google.com/site/changepointanalysis/> [accessed 2016-03-26] [WebCite Cache ID 6gHwpzPHi]
50. Pervaiz F, Pervaiz M, Abdur RN, Saif U. FluBreaks: early epidemic detection from Google flu trends. *J Med Internet Res* 2012;14(5):e125 [FREE Full text] [doi: [10.2196/jmir.2102](https://doi.org/10.2196/jmir.2102)] [Medline: [23037553](https://pubmed.ncbi.nlm.nih.gov/23037553/)]
51. Wu-Chen S. Integrating and mining virtual communities across multiple online social networks: concepts, approaches and challenges. In: *Proceedings of the Fourth International Conference on Digital Information and Communication Technology and its Applications*. 2014 Presented at: The Fourth International Conference on Digital Information and Communication Technology and its Applications; 2014; Bangkok, Thailand p. 199-204. [doi: [10.1109/DICTAP.2014.6821682](https://doi.org/10.1109/DICTAP.2014.6821682)]
52. Pew Research Center. Internet user demographics URL: <http://www.pewinternet.org/data-trend/internet-use/latest-stats/> [accessed 2016-03-26] [WebCite Cache ID 6gHwYYhq]
53. Cavazos-Rehg PA, Krauss MJ, Sowles S, Connolly S, Rosas C, Bharadwaj M, et al. A content analysis of depression-related Tweets. *Comput Human Behav* 2016 Jan 1;54:351-357. [doi: [10.1016/j.chb.2015.08.023](https://doi.org/10.1016/j.chb.2015.08.023)] [Medline: [26392678](https://pubmed.ncbi.nlm.nih.gov/26392678/)]
54. Reavley NJ, Pilkington PD. Use of Twitter to monitor attitudes toward depression and schizophrenia: an exploratory study. *PeerJ* 2014;2:e647 [FREE Full text] [doi: [10.7717/peerj.647](https://doi.org/10.7717/peerj.647)] [Medline: [25374786](https://pubmed.ncbi.nlm.nih.gov/25374786/)]
55. Chen AT, Zhu S, Conway M. What Online Communities Can Tell Us About Electronic Cigarettes and Hookah Use: A Study Using Text Mining and Visualization Techniques. *J Med Internet Res* 2015;17(9):e220 [FREE Full text] [doi: [10.2196/jmir.4517](https://doi.org/10.2196/jmir.4517)] [Medline: [26420469](https://pubmed.ncbi.nlm.nih.gov/26420469/)]
56. Krauss MJ, Sowles SJ, Moreno M, Zewdie K, Gruzca RA, Bierut LJ, et al. Hookah-Related Twitter Chatter: A Content Analysis. *Prev Chronic Dis* 2015;12:E121 [FREE Full text] [doi: [10.5888/pcd12.150140](https://doi.org/10.5888/pcd12.150140)] [Medline: [26226068](https://pubmed.ncbi.nlm.nih.gov/26226068/)]
57. Krueger EA, Young SD. Twitter: A Novel Tool for Studying the Health and Social Needs of Transgender Communities. *JMIR Ment Health* 2015;2(2):e16 [FREE Full text] [doi: [10.2196/mental.4113](https://doi.org/10.2196/mental.4113)] [Medline: [26082941](https://pubmed.ncbi.nlm.nih.gov/26082941/)]
58. Santillana M, Nguyen AT, Dredze M, Paul MJ, Nsoesie EO, Brownstein JS. Combining Search, Social Media, and Traditional Data Sources to Improve Influenza Surveillance. *PLoS Comput Biol* 2015 Oct;11(10):e1004513 [FREE Full text] [doi: [10.1371/journal.pcbi.1004513](https://doi.org/10.1371/journal.pcbi.1004513)] [Medline: [26513245](https://pubmed.ncbi.nlm.nih.gov/26513245/)]

## Abbreviations

**API:** application programming interface

**CDC:** Centers for Disease Control and Prevention

**ILI:** influenza-like illness

**ILINet:** United States Outpatient Influenza-like Illness Surveillance Network

**MCMC:** Markov Chain Monte Carlo

**MMWR:** Morbidity and Mortality Weekly Report

**PPV:** positive predictive value

*Edited by P Sullivan, G Eysenbach; submitted 23.04.16; peer-reviewed by E Slok, WC Su, M Paul, D Broniatowski; comments to author 27.07.16; revised version received 31.08.16; accepted 21.09.16; published 20.10.16.*

*Please cite as:*

*Sharpe JD, Hopkins RS, Cook RL, Striley CW*

*Evaluating Google, Twitter, and Wikipedia as Tools for Influenza Surveillance Using Bayesian Change Point Analysis: A Comparative Analysis*

*JMIR Public Health Surveill 2016;2(2):e161*

*URL: <http://publichealth.jmir.org/2016/2/e161/>*

*doi: [10.2196/publichealth.5901](https://doi.org/10.2196/publichealth.5901)*

*PMID: [27765731](https://pubmed.ncbi.nlm.nih.gov/27765731/)*

©J Danielle Sharpe, Richard S Hopkins, Robert L Cook, Catherine W Striley. Originally published in JMIR Public Health and Surveillance (<http://publichealth.jmir.org>), 20.10.2016. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Public Health and Surveillance, is properly cited. The complete bibliographic information, a link to the original publication on <http://publichealth.jmir.org>, as well as this copyright and license information must be included.

Original Paper

# “When ‘Bad’ is ‘Good’”: Identifying Personal Communication and Sentiment in Drug-Related Tweets

Raminta Daniulaityte<sup>1,2</sup>, PhD; Lu Chen<sup>2</sup>, MSc; Francois R Lamy<sup>1,2</sup>, PhD; Robert G Carlson<sup>1,2</sup>, PhD; Krishnaprasad Thirunarayan<sup>2</sup>, PhD; Amit Sheth<sup>2</sup>, PhD

<sup>1</sup>Center for Interventions, Treatment, and Addictions Research, Department of Population and Public Health Sciences, Boonshoft School of Medicine, Wright State University, Kettering, OH, United States

<sup>2</sup>The Ohio Center of Excellence in Knowledge-enabled Computing, Department of Computer Science and Engineering, Wright State University, Dayton, OH, United States

**Corresponding Author:**

Raminta Daniulaityte, PhD

Center for Interventions, Treatment, and Addictions Research

Department of Population and Public Health Sciences, Boonshoft School of Medicine

Wright State University

Kettering, OH, 45420

United States

Phone: 1 937 775 1411

Fax: 1 937 775 1419

Email: [raminta.daniulaityte@wright.edu](mailto:raminta.daniulaityte@wright.edu)

## Abstract

**Background:** To harness the full potential of social media for epidemiological surveillance of drug abuse trends, the field needs a greater level of automation in processing and analyzing social media content.

**Objectives:** The objective of the study is to describe the development of supervised machine-learning techniques for the eDrugTrends platform to automatically classify tweets by type/source of communication (personal, official/media, retail) and sentiment (positive, negative, neutral) expressed in cannabis- and synthetic cannabinoid-related tweets.

**Methods:** Tweets were collected using Twitter streaming Application Programming Interface and filtered through the eDrugTrends platform using keywords related to cannabis, marijuana edibles, marijuana concentrates, and synthetic cannabinoids. After creating coding rules and assessing intercoder reliability, a manually labeled data set (N=4000) was developed by coding several batches of randomly selected subsets of tweets extracted from the pool of 15,623,869 collected by eDrugTrends (May-November 2015). Out of 4000 tweets, 25% (1000/4000) were used to build source classifiers and 75% (3000/4000) were used for sentiment classifiers. Logistic Regression (LR), Naive Bayes (NB), and Support Vector Machines (SVM) were used to train the classifiers. Source classification (n=1000) tested Approach 1 that used short URLs, and Approach 2 where URLs were expanded and included into the bag-of-words analysis. For sentiment classification, Approach 1 used all tweets, regardless of their source/type (n=3000), while Approach 2 applied sentiment classification to personal communication tweets only (2633/3000, 88%). Multiclass and binary classification tasks were examined, and machine-learning sentiment classifier performance was compared with Valence Aware Dictionary for sEntiment Reasoning (VADER), a lexicon and rule-based method. The performance of each classifier was assessed using 5-fold cross validation that calculated average F-scores. One-tailed *t* test was used to determine if differences in F-scores were statistically significant.

**Results:** In multiclass source classification, the use of expanded URLs did not contribute to significant improvement in classifier performance (0.7972 vs 0.8102 for SVM,  $P=.19$ ). In binary classification, the identification of all source categories improved significantly when unshortened URLs were used, with personal communication tweets benefiting the most (0.8736 vs 0.8200,  $P<.001$ ). In multiclass sentiment classification Approach 1, SVM (0.6723) performed similarly to NB (0.6683) and LR (0.6703). In Approach 2, SVM (0.7062) did not differ from NB (0.6980,  $P=.13$ ) or LR ( $F=0.6931$ ,  $P=.05$ ), but it was over 40% more accurate than VADER ( $F=0.5030$ ,  $P<.001$ ). In multiclass task, improvements in sentiment classification (Approach 2 vs Approach 1) did not reach statistical significance (eg, SVM: 0.7062 vs 0.6723,  $P=.052$ ). In binary sentiment classification (positive vs negative), Approach 2 (focus on personal communication tweets only) improved classification results, compared with Approach 1, for LR (0.8752 vs 0.8516,  $P=.04$ ) and SVM (0.8800 vs 0.8557,  $P=.045$ ).

**Conclusions:** The study provides an example of the use of supervised machine learning methods to categorize cannabis- and synthetic cannabinoid-related tweets with fairly high accuracy. Use of these content analysis tools along with geographic identification capabilities developed by the eDrugTrends platform will provide powerful methods for tracking regional changes in user opinions related to cannabis and synthetic cannabinoids use over time and across different regions.

(*JMIR Public Health Surveill* 2016;2(2):e162) doi:[10.2196/publichealth.6327](https://doi.org/10.2196/publichealth.6327)

## KEYWORDS

social media; Twitter; cannabis; synthetic cannabinoids; machine learning; sentiment analysis; eDrugTrends

## Introduction

To design effective prevention, intervention, and policy measures, public health professionals require timely and reliable information on new and emerging drug use practices and trends [1-3]. There is a growing recognition that user-generated content available through Web-based and social media platforms such as Twitter, can be used as a rich data source of unsolicited and unfiltered self-disclosures of substance use and abuse behaviors. Such data could be used to complement and broaden the scope of existing illicit drug use monitoring systems by enhancing their capacity for early identification of new trends [3-6].

Twitter is a microblogging service provider and social network platform that was launched in 2006. Currently, Twitter reports 310 million monthly active users [7] that generate over 500 million tweets per day [8]. Prior research has demonstrated that Twitter can be a useful tool for infodemiology studies of very diverse public health issues [9-12]. Furthermore, the US Twitter population is young and ethnically diverse, which makes analysis of Twitter data particularly suitable for drug abuse surveillance because young adults display the highest rates of drug use behaviors [13].

Because of the high volume of data generated by Twitter users and availability of geographic information, analysis of tweets can help identify geographic and temporal trends [14-17]. The content of tweets, although brief and limited to 140 characters (with some recent relaxation of this limit), can be used to extract information on user attitudes and behaviors related to drug use [15,16,18-22]. Prior research indicates that the ability to separate personal communications from other types of communications such as official/media or retail-related tweets might help reduce the “noise” in social media research and increase the quality of the data for epidemiological surveillance [23,24]. Sentiment analysis is another approach to content analysis of social media data that seeks to understand the opinions (positive, negative, or neutral) expressed regarding selected topics.

Several prior studies used manual coding to classify cannabis, alcohol, and other drug-related tweets by sentiment [15,18,20,21] and source [15,21]. However, such studies, because they relied on manual coding, were limited to the analyses of relatively small samples of tweets. Manual coding is a labor intensive and time consuming process, and its wider application to social media data is human-resource intensive and hence slow, expensive, and difficult in particular for the purpose of identifying emerging trends in real-time. Automation of content analysis tasks would provide powerful tools to examine temporal and geographic trends not just in terms of

general tweeting activity [14-17], but also in terms of the types of communications and opinions expressed in such tweets (eg, how the opinions expressed in tweets in relation to emerging cannabis products change over time and vary across different states and regions).

Although several prior studies reported on the development of automated approaches to analyze tobacco and ecigarette-related tweet content [25,26] and to identify adverse effects associated with medical use of pharmaceutical drugs [27,28], there have been very few attempts to apply automated content analysis techniques to analyze drug abuse-related tweets [29]. This lack of research is partially related to the fact that drug-related content adds another layer of ambiguity and difficulty in the development of automated techniques because of pervasive use of slang terminology and implied meanings [30,31]. For example, the sentiment lexicon that generally conveys negative meaning in its conventional uses (eg, “bad,” “wasted,” “faded,” “fucked up”) could express positive sentiment when used in drug-related tweets that describe desired effects of getting intoxicated and high (eg, “I wanna mad amounts of blunts and let’s get faded”; “I get fucked up on this shit, I drink lean and smoke dabs every day”). For this domain-specific usage and meanings of sentiment words (where “bad” comes to mean “good,” such as in the case of being “faded” or “fucked up”), traditional approaches that use sentiment lexicons (eg, Valence Aware Dictionary for sEntiment Reasoning (VADER) [32]) may not perform well, and machine learning techniques, trained using manually coded data, could increase the accuracy of sentiment identification in drug-related tweets.

The study builds on interdisciplinary collaboration that combines drug abuse and computer science research to develop eDrugTrends, a highly scalable infoveillance platform for real-time processing of social media data related to cannabis and synthetic cannabinoid use. Development of eDrugTrends platform is based on previous research and infrastructure created by our research team, including Twitris (for analysis of Twitter data) [33-36] and PREDOSE (for analysis of Web forum data) [37-39].

The key goal of this study is to describe the development and performance of machine learning classifiers to automatically identify tweets by the source/type of communication (personal, official/media, retail) and sentiment (positive, negative, neutral) expressed in cannabis- and synthetic cannabinoid-related tweets. Because prior research identified distinct linguistic and sentiment patterns in personal communication tweets compared with tweets generated by organizational entities [15,23], the study also tests an innovative approach that integrates sentiment



and source classification to examine sentiment identification in personal communication tweets.

## Methods

### Data Collection

The eDrugTrends platform [14,15] was used to collect and filter Twitter data available through Twitter's streaming Application Programming Interface. eDrugTrends filters out non-English language tweets and uses keywords and blacklist words to extract tweets of interest. Keywords related to cannabis products (cannabis in general, marijuana edibles, marijuana concentrates) and synthetic cannabinoids were selected using prior research, media publications, and social media discussions of relevant terms [24]. To increase the accuracy of collected tweets, ambiguous slang terms (eg, blunt, spice) were combined with keywords indicating drug usage (eg, smoke/smoked/smoking). In addition, a "blacklist" of words was used to exclude collection of irrelevant tweets (eg, Emily Blunt, pumpkin spice latte) [14,15]. Performance of selected keywords was continuously monitored to identify emerging new uses, contexts, and meanings of slang terminology. The eDrugTrends platform is a real-time data collection system that initiated cannabis- and synthetic cannabinoid-related Twitter data collection in November 2014.

The Wright State University institutional review board reviewed the protocol and determined that the study meets the criteria for Human Subjects Research exemption 4 because it is limited to publicly available tweets. Tweets used as examples were modified slightly to ensure the anonymity of Twitter users who had posted them.

### Manual Coding

Manual coding was conducted to develop a labeled data set to be used as a "gold standard" for machine learning classifiers. First, 3 drug abuse researchers or "domain experts" (RD, FL, RC) conducted preliminary "open" coding [40] of several batches of 200-300 tweets to develop and refine the coding rules for source (Multimedia Appendix 1) and sentiment classification (Multimedia Appendix 2). Next, to assess intercoder reliability, a random subsample of 300 tweets was selected from a batch of 3000 tweets that were randomly extracted from eDrugTrends database of tweets collected between May and July of 2016. Reliability subsample was coded independently by the first and third authors using QDA Miner [41]. Krippendorff's Alpha statistic was used to assess intercoder reliability [42]. Coding of personal communication (K Alpha = 0.84) and media-related communication (K Alpha = 0.83) tweets had substantial agreement, while agreement was moderate for retail-related tweets (K Alpha = 0.64). Coding of positive (K Alpha = 0.69) and negative sentiment (K Alpha = 0.68) had an adequate level of agreement. However, coding of neutral/unidentified category of tweets achieved a lower level of intercoder agreement (K Alpha = 0.49), which could be explained by the fact that this category was a more amorphous and eclectic group.

Development of the manually labeled data set involved several phases of coding conducted by the first and third authors. To obtain a more balanced dataset, less common categories (eg,

negative or retail-related tweets) were purposefully oversampled (for more details, see Multimedia Appendix 3). Oversampling of underrepresented categories is important in order to obtain a more balanced data set for development of machine learning classifiers, given that significant under sampling of a certain category in the training data can directly impact the quality of classification [26]. To reach a sample size of 4000 tweets for the manually labeled data set for machine learning, more than 8000 tweets were manually reviewed and filtered using QDA Miner [41]. The tweets for manual coding were extracted from the pool of 15,623,869 tweets that were collected by eDrugTrends between May and November 2015.

The sample of 4,000 manually labeled tweets was split into two subsamples—1000 were used to train source classifier, and 3000 were allocated for sentiment classification. Information on the manually labeled tweet numbers by category for each subsample is provided in Multimedia Appendix 4.

### Machine Learning

Because the study aimed to integrate source and sentiment classification by focusing on sentiment in personal communication tweets only, source classification can be seen as a preprocessing step that is done before sentiment classification. First, 1000 tweets were used to train a source classifier (Multimedia Appendix 4). Next, for the remaining 3000 tweets (Multimedia Appendix 4), the source classifier is applied to filter out the media- and retail-related tweets, and then train the sentiment classifiers using only the personal communication tweets.

### Source Classification Models

Development of source classifiers focused only on tweets with URLs. Because all media- and retail-related tweets contained URLs, tweets without URLs could be automatically classified as belonging to the personal communication category. To select 1000 tweets with URLs for source classifier, approximately equal numbers of tweets were randomly sampled from each category—330 official/media-related, 340 retail-related, and 330 tweets that contain URLs from personal communication.

Summary information about the machine learning classification models used in the study is presented in Textbox 1. Source classification tested 2 approaches: Approach 1 used short URLs as they appear in tweets, and Approach 2 expanded URLs to their original version and used unigrams and bigrams obtained from unshortened URLs as features in machine learning (Textbox 1 A). Twitter automatically shortens all links to save character space [43], and such shortened links typically do not contain identifiable words. In contrast, expanded URLs frequently contain useful information that could help improve tweet classification accuracy. Examples of commonly occulting words identified in expanded URLs are presented in Multimedia Appendix 5.

First, performance of source classifiers was assessed for multiclass classification (media, retail, personal). Next, the best performing machine learning algorithm in multiclass classification was selected to assess 3 binary classification tasks: (1) media versus the remaining tweets, (2) retail versus the

remaining tweets, and (3) personal communication tweets versus the remaining tweets (Textbox 1 A).

### Sentiment Classification Models

Sentiment classification tested 2 approaches: Approach 1 applied sentiment classification to all tweets, regardless of their source/type, using all 3000 manually labeled tweets (1292 positive, 921 negative, 787 neutral/unidentifiable), and Approach 2 applied sentiment classification to tweets identified as personal communications only, excluding retail and media-related tweets. For this approach, the sample of 3000 tweets was first processed using the best performing source classifier (developed for this study) to identify personal communication tweets, which resulted in a sample of 2633 tweets (Textbox 1 B). The sample of 2633 tweets contained 1157 that were manually labeled as positive, 850 negative, and 626 neutral/unidentifiable. (Note that these numbers are different from the information presented in Multimedia Appendix 4 because extraction of 2633 personal communication tweets was performed using source classifier, while Multimedia Appendix 4 information is based on manual coding).

Performance of sentiment classifiers was examined for multiclass (positive, negative, neutral) and for binary classification tasks. Binary classification focused on positive versus negative tweets to examine how well sentiment classifiers performed on reliable categories (as determined by reliability assessment), excluding neutral/unidentifiable group that reached a low level of agreement among human coders. To test Approach 1 (all tweets, regardless of source/type), binary classification used a data set of 2213 tweets that was obtained after removing 787 neutral tweets from the sample of 3000. To test Approach 2 (personal communication tweets only), binary classification used a dataset of 2007 tweets that was obtained after removing 626 neutral/unidentifiable tweets from the sample of 2633 (Textbox 1 B).

In addition, the study used a lexicon and rule-based method VADER that was developed for the analysis of social media texts [32] to classify manually labeled tweet sample allocated for sentiment analysis (N=3000). VADER performance in classifying manually annotated tweets was compared with the accuracy of machine learning classifiers using a one-tailed *t* test statistic.

**Textbox 1.** Summary information on classification models tested for tweet classification by source/type and sentiment.

#### A. Classification by source/type

**Approach 1:** using all tweets, regardless of their source/type

- Multiclass classification [logistic regression (LR), naive bayes (NB), support vector machines (SVM)]:
  - Personal versus media versus retail (n=1000)
- Binary classification (using classifier that showed the best results in multiclass classification):
  - Personal versus the rest (n=1000)
  - Retail versus the rest (n=1000)
  - Media versus the rest (n=1000)

**Approach 2:** using expanded URLs

- Multiclass classification (LR, NB, SVM):
  - Personal versus media versus retail (N=1000)
- Binary Classification (using classifier that showed the best results in multiclass classification):
  - Personal versus the rest (n=1000)
  - Retail versus the rest (n=1000)
  - Media versus the rest (n=1000)

#### B. Classification by sentiment

**Approach 1:** using all tweets, regardless of their source/type

- Multiclass classification (LR, NB, SVM):
  - Positive versus negative versus neutral/unknown (N=3000)
- Binary Classification (LR, NB, SVM):
  - Positive versus negative (N=2213; neutral excluded)

**Approach 2:** using personal communication tweets only

- Multiclass classification (LR, NB, SVM):
  - Positive versus negative versus neutral/unknown (N=2633)
- Binary Classification (LR, NB, SVM):
  - Positive versus negative (N=2007; neutral/unknown excluded)

## Building and Assessment of Machine Learning Classifiers

To build classifiers, the tweets were tokenized and all words were processed to convert uppercase letters to lowercase. Because prior research suggests that stop words and complete forms of words can be useful sentiment indicators, particularly in brief texts such as tweets, stop words were retained, and no stemming was applied [44-46]. Next, all the unigrams and bigrams were collected and chi-square test was applied to select the top 500 unigrams and bigrams with highest chi-square scores as features [47]. For each feature  $t$  ( $i$ ), its tf-idf score was calculated in a tweet  $d$  ( $j$ ) as  $w(i,j) = \text{tf}(i,j) \times \text{idf}(i)$ . Term frequency  $\text{tf}(i,j)$  is the number of times feature  $t$  ( $i$ ) occurs in tweet  $d$  ( $j$ ). Inverse document frequency is calculated as  $\text{idf}(i) = \log(N/\text{df}(i))$ , where  $N$  is the total number of tweets in the dataset, and  $\text{df}(i)$  is the number of tweets in which feature  $t$  ( $i$ ) occurs. Each tweet is represented as a feature vector, and each entry of the vector is the tf-idf score of that feature in the tweet. Three machine learning classification techniques were tested for each classification model/approach: Logistic Regression (LR), Naive Bayes (NB), and Support Vector Machines (SVM). All three are commonly used classification algorithms that are known to achieve good results on text classification tasks [25,26,48,49].

The performance of each classifier was assessed by 5-fold cross validation, which is a commonly used method for the evaluation of classification algorithms that diminishes the bias in the estimation of classifier performance [50]. This approach uses the entire dataset for both training and testing, and is especially useful when the manually labeled data set is relatively small. In 5-fold cross-validation, the manually labeled data set is

randomly partitioned into 5 equal-sized subsets. The cross-validation process is then repeated 5 times (the folds). Each time, a single subset is retained as the validation data for testing the model, and the remaining 4 subsamples are used as training data. The 5 results from the folds are then averaged to produce a single estimation. The study reports the average of the precision, recall, and F-scores calculated by the system on different folds. Precision is defined as the number of correctly classified positive examples divided by the number of examples labeled by the system as positive. Recall, also referred to as sensitivity, is defined as the number of correctly classified positive examples divided by the number of positive examples in the manually coded data. An F-score is a combination (harmonic mean) of precision and recall measures [51]. One-tailed  $t$  test statistic was used to determine which classifiers performed significantly better ( $P < .05$ ).

## Results

### Source Classification

Source classification (Approach 1) that used short URLs demonstrated good performance (Table 1 A). SVM algorithm applied to multiclass classification task achieved a macro average F-score of 0.7972, which was not significantly higher compared with LR ( $P = .09$ ) or NB ( $P = .27$ ) performance (Table 1 A). Table 1 B shows the performance of source classifier that used expanded URLs when applied to multiclass classification task. SVM showed slightly better improvement in performance in multiclass classification, compared with NB and LR algorithms, reaching 0.8141 precision, 0.8119 recall, and an F-score of 0.8102. However, these differences did not reach a level of statistical significance (Table 1 C).

**Table 1.** Performance of multiclass source classifiers.

Algorithm	Precision	Recall	F-Score
<b>A. Approach 1, using short URLs (N=1000)</b>			
LR <sup>a</sup>	0.8007	0.7946	0.7938
NB <sup>b</sup>	0.8023	0.7926	0.7936
SVM <sup>c</sup>	0.8059	0.7976	0.7972
<b>B. Approach 2, using the unshortened URLs as features (N=1000)</b>			
LR	0.8062	0.8026	0.8013
NB	0.8005	0.7972	0.7953
SVM	0.8141	0.8119	0.8102
<b>C. P values calculated using t test to assess statistical significance of differences in classifier performance (F-scores)</b>			
Approach 1	SVM vs LR, $P = .09$ ; SVM vs NB, $P = .27$		
Approach 2	SVM vs LR, $P = .13$ ; SVM vs NB, $P = .10$		
Approach 1 vs 2	LR1 vs LR2, $P = .19$ ; NB1 vs NB2, $P = .47$ ; SVM1 vs SVM2, $P = .19$		

<sup>a</sup>LR: logistic regression.

<sup>b</sup>NB: naive bayes.

<sup>c</sup>SVM: support vector machines.

**Table 2.** Performance of SVM source classifiers on binary classification for each source category.

Type of classification	Precision	Recall	F-Score
<b>A. Approach 1, using short URLs (N=1000)</b>			
Media	0.8873	0.8278	0.8477
Retail	0.8723	0.7913	0.8117
Personal	0.8755	0.7976	0.8200
<b>B. Approach 2, using unshortened URLs (N=1000)</b>			
Media	0.8958	0.8639	0.8769
Retail	0.8881	0.8155	0.8357
Personal	0.9020	0.8572	0.8736
<b>C. P values calculated using t test to assess statistical significance of differences in classifier performance (F-scores)</b>			
Approach 1	Personal vs Media, $P=.094$ ; Personal vs Retail, $P=.27$ , Media vs Retail, $P=.07$		
Approach 2	Personal vs. Media, $P=.38$ ; Personal vs Retail, $P=.03^a$ ; Media vs Retail, $P=.01^a$		
Approach 1 vs 2	Personal1 vs. Personal2, $P=.001^a$ ; Retail1 vs Retail2, $P=.004^a$ ; Media1 vs Media2, $P=.049^a$		

<sup>a</sup>Values that show statistically significant differences.

Performance of both source classification approaches was also assessed on binary classification tasks. Because SVM showed slightly better performance in multiclass classification than NB or LR (although not statistically significant), it was selected for evaluation on 3 binary classification tasks using the 1000 tweets: (1) media-related tweets versus the rest of tweets, (2) retail-related tweets versus the rest of tweets, and (3) personal tweets versus the rest of tweets (Table 2). When using short URLs for binary classification task, identification of media-related tweets showed slightly better precision, recall, and overall F-scores compared with identification of retail or personal communication tweets (Table 2 A), although these differences were not statistically significant (Table 2 C). The identification of all 3 source categories benefited significantly when unshortened URLs were used as features in classification. Improvements in F-scores between Approaches 1 and 2 were significant for all 3 categories (Table 2 C). Identification of the personal communication tweets benefited the most reaching 0.9020 precision, 0.8572 recall, and an F-score of 0.8736, compared with an F-score of 0.8200 when using short URLs ( $P<.001$ ). Furthermore, when Approach 2 was used, identification of media and personal communication tweets showed significantly higher F-scores compared with retail-related tweet identification (Table 2 C).

### Sentiment Classification

For general sentiment classification approach that classified all 3000 tweets regardless of their source, SVM results showed better precision (0.7147) than other machine learning classifiers, but LR achieved better recall (0.6763) (Table 3 A). In overall F-scores, SVM achieved slightly better results ( $F=0.6723$ ) than other machine learning classifiers, but the differences were not

statistically significant (Table 3 C). However, all 3 machine-learning algorithms achieved better results than the lexicon and rule based method VADER. Compared with VADER ( $F=0.5116$ ), SVM performance was over 30% better, and the difference was statistically significant at  $P<.001$  (Table 3 C).

Before sentiment classification Approach 2 could be applied, the sample of 3000 tweets had to be processed to extract personal communication tweets. Because the SVM source classifier with unshortened URLs showed better performance than other classifiers (Table 2), it was used to identify the personal communication tweets (2633) from the sample of 3000. Table 3 B shows evaluation of sentiment classification of personal communication tweets. Compared with Approach 1 (Table 3 A), multiclass sentiment classification of personal communication tweets (Approach 2) showed approximately 3% improvement for NB, 4% improvement for LR, and 5% for SVM classifier, although these increases did not reach a level of statistical significance (Table 3 C). The NB classifier achieved the greatest precision (0.7539), but SVM showed the highest recall scores (0.7021). Overall, the SVM classifier demonstrated slightly better performance than the other 2 machine learning classifiers by achieving an F-score of 0.7062, which was significantly greater compared with LR and NB, but these difference did not reach statistical significance. All 3 machine-learning classifiers achieved better accuracy than VADER. The F-score of SVM was over 40% greater in comparison to VADER performance, and the difference was statistically significant at  $P<.001$  (Table 3 C). The most discriminative unigram and bigram features reflect thematic categories pertinent to each source category (Multimedia Appendix 6).

**Table 3.** Performance of multiclass sentiment classifiers (positive, negative, neutral).

Algorithm	Precision	Recall	F-Score
<b>A. Approach 1, including all tweets regardless of the source (N=3000)</b>			
LR <sup>a</sup>	0.7047	0.6763	0.6703
NB <sup>b</sup>	0.7101	0.6693	0.6683
SVM <sup>c</sup>	0.7147	0.6691	0.6723
VADER <sup>d</sup>	0.5213	0.5261	0.5116
<b>B. Approach 2, including personal communication tweets only (N=2633)</b>			
LR	0.7145	0.6996	0.6931
NB	0.7539	0.6914	0.6980
SVM	0.7442	0.7021	0.7062
VADER	0.5153	0.5211	0.5030
<b>C. P values calculated using t test to assess statistical significance of differences in classifier performance (F-scores)</b>			
Approach 1	SVM vs LR, $P=.38$ ; SVM vs NB, $P=.23$ ; SVM vs VADER, $P<.001$ <sup>e</sup>		
Approach 2	SVM vs LR, $P=.05$ ; SVM vs NB, $P=.13$ ; SVM vs VADER, $P<.001$ <sup>e</sup>		
Approach 1 vs 2	LR1 vs LR2, $P=.08$ ; NB1 vs NB2, $P=.06$ ; SVM1 vs SVM2, $P=.052$		

<sup>a</sup>LR: logistic regression.

<sup>b</sup>NB: naive bayes.

<sup>c</sup>SVM: support vector machines.

<sup>d</sup>VADER: Valence Aware Dictionary for sEntiment Reasoning.

<sup>e</sup>Values that show statistically significant differences.

**Table 4.** Performance of binary sentiment classifiers (positive vs negative).

Algorithm	Precision	Recall	F-Score
<b>A. Approach 1, including all tweets regardless of the source (N=2213)</b>			
LR <sup>a</sup>	0.8700	0.8495	0.8516
NB <sup>b</sup>	0.8797	0.8491	0.8540
SVM <sup>c</sup>	0.8803	0.8513	0.8557
<b>B. Approach 2, personal communication tweets only (N=2007)</b>			
LR	0.8878	0.8728	0.8752
NB	0.8892	0.8629	0.8666
SVM	0.8964	0.8757	0.8800
<b>C. P values calculated using t test to assess statistical significance of differences in classifier performance (F-scores)</b>			
Approach 1	SVM vs LR, $P=.20$ ; SVM vs NB, $P=.36$ ;		
Approach 2	SVM vs LR, $P=.20$ ; SVM vs NB, $P=.003$ <sup>d</sup> ;		
Approach 1 vs 2	LR1 vs LR2, $P=.04$ <sup>d</sup> ; NB1 vs NB2, $P=.13$ ; SVM1 vs SVM2, $P=.045$ <sup>d</sup>		

<sup>a</sup>LR: logistic regression.

<sup>b</sup>NB: naive bayes.

<sup>c</sup>SVM: support vector machines.

<sup>d</sup>Values that show statistically significant differences.

As shown in Table 4 A, for binary sentiment classification (Approach 1), the SVM classifier showed the best precision and recall scores. The SVM algorithm achieved an F-score of 0.8557, which was slightly higher than LR and NB, although

the differences were not statistically significant (Table 4 C). When sentiment classification was performed on personal communication tweets only (Table 4 B), LR and SVM performance showed statistically significant improvement in



comparison to Approach 1 binary classification task (Table 4 C). The SVM classifier achieved high precision and recall (both of which approached 90%), and an F-score of 0.8800, which was significantly greater in comparison to NB, but not significantly different from LR (Table 4 C). Results of binary classification tasks were not compared with VADER, because the latter still classifies tweets into 3 categories assigning a tweet to a neutral category when it cannot find any sentiment words/patterns.

The most discriminative unigram and bigram features that were identified by chi-square test reflect thematic groups as pertinent to sentiment categories: “want,” “love,” “need” for positive, in contrast to “don’t,” “shit,” “fake” for negative tweets (Multimedia Appendix 7). Our sentiment classifier tended to incorrectly classify tweets that expressed an opposing opinion to negative thoughts or actions related to cannabis use or its legalization. For example, the following tweets were classified as negative by our classifier, although manual coding identified them as conveying positive views toward cannabis: “@GovChristie very ignorant to not see the value of cannabis”; “I think it's ridiculous professional athletes get penalized for smoking a joint...”. Humorous and sarcastic tweets were also more difficult to classify correctly by our classifier. For example, the following tweet was coded by domain experts as conveying a positive attitude toward marijuana, but was coded as negative by our machine learning classifier: “Marijuana - side effects may include being happy and consumption of fast food.”

## Discussion

### Principal Findings

The results of this study provide an example of the use of supervised machine learning methods to categorize cannabis- and synthetic cannabinoid-related content on Twitter with fairly high accuracy. To classify tweets by source/type of communication, an SVM algorithm that used expanded URLs produced the best results, in particular as demonstrated by binary classification tasks. For sentiment classification, the SVM algorithm that focused on “personal communication” tweets, in particular classifying positive versus negative tweets only, performed better than a more general approach that included all tweets regardless of the source.

Integration of the 2 dimensions of content analysis tasks—identification of type of communication and sentiment—represents a novel approach. Identification of sentiment in user-generated tweets (personal communications) carries greater relevance for drug abuse epidemiology research than an approach that does not separate personal from media- and retail-related tweets. Use of these content analysis tools along with geographic identification features currently functional in the eDrugTrends platform [14] will provide powerful methods for tracking regional changes in user sentiments related to cannabis and synthetic cannabinoids use over time and across different states or regions.

Overall, our machine learning methods for sentiment classification demonstrated substantially better performance than the lexicon and rule-based method VADER [32]. Prior

research has shown that VADER method can achieve an F-score of 0.96 in identifying sentiment when applied to “general” tweets. It is noteworthy that VADER accuracy in classifying tweets in drug use-related domain (where negative words sometime can convey positive and desired experiences) was substantially lower (F=0.51). The accuracy of SVM multiclass sentiment classifier that focused on personal communication tweets only was 40% better in comparison to VADER performance, and the difference was statistically significant at  $P<.001$ .

Our study demonstrates that content analysis and manual coding of drug-related tweets is not an easy task even for human coders with substantial experience in drug abuse research and qualitative content analysis. This is consistent with prior studies that have reported high level of ambiguity and lack of context as complicating factors in content analysis of tweets [52]. Although our study demonstrates strong performance of machine learning classifiers for automatic classification of tweet content, manual coding will remain an important method necessary for exploration of new domains and improvement of existing automated classification techniques to reflect changes in drug use practices and/or slang terminology. Our experiences developing the labeled data set emphasize the importance of: (1) revealing ambiguities and difficulties encountered when conducting manual coding, and (2) using appropriate metrics to assess intercoder reliability [42].

### Limitations

One of the limitations of our study is that we did not include development of machine learning classification methods to identify relevant and irrelevant tweets (eg, cases were “spice” may refer not to synthetic cannabinoids but to food seasoning). Relevance of extracted data was monitored using appropriate keyword combinations and blacklisted words [15]. We also note the limitations in relation to our ability to identify neutral tweets because they were grouped together with the “unidentifiable” or “difficult to classify” tweets. Until better methods are developed, our future applications of eDrugTrends sentiment analysis tools will take into consideration that neutral/unidentifiable group is a nonreliable category, and will focus on drawing conclusions about positive/negative sentiment tweets only.

Future research will assess performance of these techniques to analyze tweets mentioning other drugs of abuse and will also extend them to automate extraction of more detailed thematic information from drug-related tweets. In addition, because many tweets contain visual information to convey meaning, machine learning-based image classification would add an additional dimension and improve the accuracy of overall tweet content classification. In the future, we will examine the feasibility of separating true neutral tweets from unidentifiable group to improve sentiment analysis.

### Conclusions

This is one of the first studies to report successful development of automated content classification tools to analyze recreational drug use-related tweets. These tools, as a part of eDrugTrends platform, will help advance the field’s technological and

methodological capabilities to harness social media sources for drug abuse surveillance research. Our future deployment of the eDrugTrends platform will generate data on emerging regional and temporal trends and inform more timely interventions and policy responses to changes in cannabis and synthetic cannabinoid use practices.

---

### Acknowledgments

This study was supported by the National Institute on Drug Abuse (NIDA), Grant No. R01 DA039454 (Daniulaityte, PI; Sheth, PI). The funding source had no further role in the study design, in the collection, analysis, and interpretation of the data, in the writing of the report, or in the decision to submit the paper for publication.

---

### Conflicts of Interest

None declared.

---

### Multimedia Appendix 1

Source classification: coding guidelines used to manually annotate tweets as personal, retail-, and media-related communications.

[[PDF File \(Adobe PDF File\), 24KB - publichealth\\_v2i2e162\\_app1.pdf](#) ]

---

### Multimedia Appendix 2

Sentiment classification: coding guidelines used to manually annotate tweets as expressing positive, negative, or neutral/unidentifiable sentiment.

[[PDF File \(Adobe PDF File\), 33KB - publichealth\\_v2i2e162\\_app2.pdf](#) ]

---

### Multimedia Appendix 3

Description of the development of manually labeled data set.

[[PDF File \(Adobe PDF File\), 16KB - publichealth\\_v2i2e162\\_app3.pdf](#) ]

---

### Multimedia Appendix 4

Information about the manually labeled tweets included in subsets to train source and sentiment classifiers.

[[PDF File \(Adobe PDF File\), 31KB - publichealth\\_v2i2e162\\_app4.pdf](#) ]

---

### Multimedia Appendix 5

Commonly occurring words in unshortened URLs by source/type category.

[[PDF File \(Adobe PDF File\), 22KB - publichealth\\_v2i2e162\\_app5.pdf](#) ]

---

### Multimedia Appendix 6

Top 10 most discriminative unigram and bigram features for source classification.

[[PDF File \(Adobe PDF File\), 19KB - publichealth\\_v2i2e162\\_app6.pdf](#) ]

---

### Multimedia Appendix 7

Top 10 most discriminative unigram and bigram features for sentiment classification.

[[PDF File \(Adobe PDF File\), 19KB - publichealth\\_v2i2e162\\_app7.pdf](#) ]

---

### References

1. Richardson J, Fendrich M, Johnson TP. Neighborhood effects on drug reporting. *Addiction* 2003;98:1705-1711. [Medline: [14651502](#)]
2. Sloboda Z. *Epidemiology of Drug Abuse*. New York: Springer; 2005.
3. Mounteney J, Haugland S. Earlier warning: a multi-indicator approach to monitoring trends in the illicit use of medicines. *Int J Drug Policy* 2009;20:161-169. [doi: [10.1016/j.drugpo.2007.09.006](#)] [Medline: [18032012](#)]

4. Corazza O, Assi S, Simonato P, Corkery J, Bersani FS, Demetrovics Z, et al. Promoting innovation and excellence to face the rapid diffusion of novel psychoactive substances in the EU: the outcomes of the ReDNet project. *Hum Psychopharmacol* 2013;28:317-323. [doi: [10.1002/hup.2299](https://doi.org/10.1002/hup.2299)] [Medline: [23881879](#)]
5. Davey Z, Schifano F, Corazza O, Deluca P, Psychonaut Web Mapping Group. e-Psychonauts: conducting research in online drug forum communities. *J Ment Health* 2012;21:386-394. [doi: [10.3109/09638237.2012.682265](https://doi.org/10.3109/09638237.2012.682265)] [Medline: [22823094](#)]
6. Miller PG, S nderlund AL. Using the internet to research hidden populations of illicit drug users: a review. *Addiction* 2010;105:1557-1567. [doi: [10.1111/j.1360-0443.2010.02992.x](https://doi.org/10.1111/j.1360-0443.2010.02992.x)] [Medline: [20626378](#)]
7. Twitter. 2016. About Company URL: <https://about.twitter.com/company> [accessed 2016-10-13] [WebCite Cache ID [6EjmJkf1](#)]
8. Internet-Live-Stats. 2016. Twitter Usage Statistics URL: <http://www.internetlivestats.com/twitter-statistics/> [accessed 2016-10-13] [WebCite Cache ID [6Ejp8QRn](#)]
9. Chew C, Eysenbach G. Pandemics in the age of Twitter: content analysis of Tweets during the 2009 H1N1 outbreak. *PLoS One* 2010;5:e14118 [FREE Full text] [doi: [10.1371/journal.pone.0014118](https://doi.org/10.1371/journal.pone.0014118)] [Medline: [21124761](#)]
10. Prieto VM, Matos S,  lvarez M, Casheda F, Oliveira JL. Twitter: a good place to detect health conditions. *PLoS One* 2014;9:e86191 [FREE Full text] [doi: [10.1371/journal.pone.0086191](https://doi.org/10.1371/journal.pone.0086191)] [Medline: [24489699](#)]
11. Weeg C, Schwartz HA, Hill S, Merchant RM, Arango C, Ungar L. Using Twitter to measure public discussion of diseases: a case study. *JMIR Public Health Surveill* 2015;1:e6 [FREE Full text] [doi: [10.2196/publichealth.3953](https://doi.org/10.2196/publichealth.3953)] [Medline: [26925459](#)]
12. 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:e11 [FREE Full text] [doi: [10.2196/jmir.1157](https://doi.org/10.2196/jmir.1157)] [Medline: [19329408](#)]
13. Substance Abuse and Mental Health Services Administration. Results from the 2013 National Survey on Drug UseHealth: National Findings. Rockville, MD: Substance Abuse and Mental Health Services Administration; 2014.
14. Daniulaityte R, Nahhas RW, Wijeratne S, Carlson RG, Lamy FR, Martins SS, et al. "Time for dabs": analyzing Twitter data on marijuana concentrates across the U.S. *Drug Alcohol Depend* 2015;155:307-311. [doi: [10.1016/j.drugalcdep.2015.07.1199](https://doi.org/10.1016/j.drugalcdep.2015.07.1199)] [Medline: [26338481](#)]
15. Lamy FR, Daniulaityte R, Sheth A, Nahhas RW, Martins SS, Boyer EW, et al. "Those edibles hit hard": exploration of Twitter data on cannabis edibles in the U.S. *Drug Alcohol Depend* 2016;164:64-70. [doi: [10.1016/j.drugalcdep.2016.04.029](https://doi.org/10.1016/j.drugalcdep.2016.04.029)] [Medline: [27185160](#)]
16. Hanson CL, Burton SH, Giraud-Carrier C, West JH, Barnes MD, Hansen B. Tweaking and tweeting: exploring Twitter for nonmedical use of a psychostimulant drug (Adderall) among college students. *J Med Internet Res* 2013;15:e62 [FREE Full text] [doi: [10.2196/jmir.2503](https://doi.org/10.2196/jmir.2503)] [Medline: [23594933](#)]
17. West JH, Hall PC, Hanson CL, Prier K, Giraud-Carrier C, Neeley ES, et al. Temporal variability of problem drinking on Twitter. *OJPM* 2012;02:43-48. [doi: [10.4236/ojpm.2012.21007](https://doi.org/10.4236/ojpm.2012.21007)]
18. Shutler L, Nelson LS, Portelli I, Blachford C, Perrone J. Drug use in the Twittersphere: a qualitative contextual analysis of tweets about prescription drugs. *J Addict Dis* 2015;34:303-310. [doi: [10.1080/10550887.2015.1074505](https://doi.org/10.1080/10550887.2015.1074505)] [Medline: [26364675](#)]
19. Thompson L, Rivara FP, Whitehill JM. Prevalence of marijuana-related traffic on Twitter, 2012-2013: a content analysis. *Cyberpsychol Behav Soc Netw* 2015;18:311-319 [FREE Full text] [doi: [10.1089/cyber.2014.0620](https://doi.org/10.1089/cyber.2014.0620)] [Medline: [26075917](#)]
20. Cavazos-Rehg PA, Krauss M, Fisher SL, Salyer P, Gruzca RA, Bierut LJ. Twitter chatter about marijuana. *J Adolesc Health* 2015;56:139-145. [doi: [10.1016/j.jadohealth.2014.10.270](https://doi.org/10.1016/j.jadohealth.2014.10.270)] [Medline: [25620299](#)]
21. Cavazos-Rehg PA, Krauss MJ, Sowles SJ, Bierut LJ. "Hey Everyone, I'm Drunk." an evaluation of drinking-related Twitter chatter. *J Stud Alcohol Drugs* 2015;76:635-643 [FREE Full text] [Medline: [26098041](#)]
22. Krauss MJ, Gruzca RA, Bierut LJ, Cavazos-Rehg PA. "Get drunk. Smoke weed. Have fun.": A content analysis of tweets about marijuana and alcohol. *Am J Health Promot* 2015. [doi: [10.4278/ajhp.150205-QUAL-708](https://doi.org/10.4278/ajhp.150205-QUAL-708)] [Medline: [26559715](#)]
23. Yin P, Ram N, Lee W, Tucker C, Khandelwal S, Salathe M. Two sides of a coin: separating personal communication and public dissemination accounts in Twitter. In: *Advances in Knowledge Discovery and Data Mining.*: Springer International Publishing; 2014 Presented at: 18th Pacific-Asia Conference on Knowledge Discovery and Data Mining; May 13-16, 2014; Tainan, Taiwan p. 163-175. [doi: [10.1007/978-3-319-06608-0\\_14](https://doi.org/10.1007/978-3-319-06608-0_14)]
24. Kim Y, Huang J, Emery S. Garbage in, garbage out: data collection, quality assessment and reporting standards for social media data use in health research, infodemiology and digital disease detection. *J Med Internet Res* 2016;18:e41 [FREE Full text] [doi: [10.2196/jmir.4738](https://doi.org/10.2196/jmir.4738)] [Medline: [26920122](#)]
25. Mysl n 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:e174 [FREE Full text] [doi: [10.2196/jmir.2534](https://doi.org/10.2196/jmir.2534)] [Medline: [23989137](#)]
26. Cole-Lewis H, Varghese A, Sanders A, Schwarz M, Pugatch J, Augustson E. Assessing electronic cigarette-related tweets for sentiment and content using supervised machine learning. *J Med Internet Res* 2015;17:e208 [FREE Full text] [doi: [10.2196/jmir.4392](https://doi.org/10.2196/jmir.4392)] [Medline: [26307512](#)]
27. Jiang K, Zheng Y. Mining Twitter data for potential drug effects. In: *Advanced Data Mining and Applications.*: Springer; 2013 Presented at: 9th International Conference ADMA; December 14-16, 2013; Hangzhou, China. [doi: [10.1007/978-3-642-53914-5\\_37](https://doi.org/10.1007/978-3-642-53914-5_37)]

28. Plachouras V, Keidner J, Garrow A. Quantifying self-reported adverse drug events on Twitter: signal and topic analysis. In: Proceedings of the 7th 2016 International Conference on Social Media & Society. 2016 Presented at: International Conference on Social Media & Society; July 11-13, 2016; London. [doi: [10.1145/2930971.2930977](https://doi.org/10.1145/2930971.2930977)]
29. Sarker A, O'Connor K, Ginn R, Scotch M, Smith K, Malone D, et al. Social media mining for toxicovigilance: automatic monitoring of prescription medication abuse from Twitter. *Drug Saf* 2016;39:231-240 [FREE Full text] [doi: [10.1007/s40264-015-0379-4](https://doi.org/10.1007/s40264-015-0379-4)] [Medline: [26748505](https://pubmed.ncbi.nlm.nih.gov/26748505/)]
30. Katsuki T, Mackey TK, Cuomo R. Establishing a link between prescription drug abuse and illicit online pharmacies: analysis of Twitter data. *J Med Internet Res* 2015;17:e280 [FREE Full text] [doi: [10.2196/jmir.5144](https://doi.org/10.2196/jmir.5144)] [Medline: [26677966](https://pubmed.ncbi.nlm.nih.gov/26677966/)]
31. Na J, Kyaing WYM. Sentiment analysis of user-generated content on drug review websites. *Journal of Information Science Theory and Practice* 2015;3:6-23. [doi: [10.1633/JISTaP.2015.3.1.1](https://doi.org/10.1633/JISTaP.2015.3.1.1)]
32. Hutto C, Gilbert E. VADER: a parsimonious rule-based model for sentiment analysis of social media text. In: Proceedings of the Eighth International Association for the Advancement of Artificial Intelligence Conference on Weblogs and Social Media. 2014 Presented at: International AAI Conference; 2014; Ann Arbor, MI URL: <http://www.aaai.org/>
33. Jadhav A, Purohit H, Kapanipathi P, Anantharam P, Ranabahu A, Nguyen V, et al. Twitris 2.0: semantically empowered system for understanding perceptions from social data. 2010 Presented at: ISWC 2010 Semantic Web Application Challenge; 2010; knoesis URL: <http://corescholar.libraries.wright.edu/cgi/viewcontent.cgi?article=1253&context=knoesis>
34. Nagarajan M, Gomadam K, Sheth A, Ranabahu A, Mutharaju R, Jadhav A. Spatio-temporal-thematic analysis of citizen sensor data: challenges and experiences. In: Proceedings of the Web Information Systems Engineering - WISE 2009. Berlin: Springer; 2009 Presented at: 10th International Conference WISE; October 5-7, 2009; Poznen, Poland p. 539-553.
35. Sheth A. Citizen sensing, social signals, and enriching human experience. *IEEE Internet Comput* 2009;13:87-92. [doi: [10.1109/MIC.2009.77](https://doi.org/10.1109/MIC.2009.77)]
36. Chen L, Wang W, Sheth A. Are Twitter users equal in predicting elections? A study of user groups in predicting 2012 US Republican Presidential Primaries. In: *Social Informatics*. Berlin: Springer; 2012:379-392.
37. Cameron D, Smith GA, Daniulaityte R, Sheth AP, Dave D, Chen L, et al. PREDOSE: a semantic web platform for drug abuse epidemiology using social media. *J Biomed Inform* 2013;46:985-997 [FREE Full text] [doi: [10.1016/j.jbi.2013.07.007](https://doi.org/10.1016/j.jbi.2013.07.007)] [Medline: [23892295](https://pubmed.ncbi.nlm.nih.gov/23892295/)]
38. Daniulaityte R, Carlson R, Falck R, Cameron D, Perera S, Chen L, et al. "I just wanted to tell you that loperamide WILL WORK": a web-based study of extra-medical use of loperamide. *Drug Alcohol Depend* 2013;130:241-244 [FREE Full text] [doi: [10.1016/j.drugalcdep.2012.11.003](https://doi.org/10.1016/j.drugalcdep.2012.11.003)] [Medline: [23201175](https://pubmed.ncbi.nlm.nih.gov/23201175/)]
39. Daniulaityte R, Carlson R, Brigham G, Cameron D, Sheth A. "Sub is a weird drug:" a web-based study of lay attitudes about use of buprenorphine to self-treat opioid withdrawal symptoms. *Am J Addict* 2015;24:403-409. [doi: [10.1111/ajad.12213](https://doi.org/10.1111/ajad.12213)] [Medline: [26009867](https://pubmed.ncbi.nlm.nih.gov/26009867/)]
40. Moltu C, Stefansen J, Svisdahl M, Veseth M. Negotiating the coresearcher mandate - service users' experiences of doing collaborative research on mental health. *Disabil Rehabil* 2012;34:1608-1616. [doi: [10.3109/09638288.2012.656792](https://doi.org/10.3109/09638288.2012.656792)] [Medline: [22489612](https://pubmed.ncbi.nlm.nih.gov/22489612/)]
41. Provalis Research. QDA Miner. Montreal, Canada: Provalis; 2014. URL: <https://provalisresearch.com/products/qualitative-data-analysis-software/> [accessed 2016-10-13] [WebCite Cache ID 6lF4wDyHY]
42. Krippendorff K. *Content Analysis: An Introduction to its Methodology*. Los Angeles: SAGE; 2012.
43. Twitter. 2016. Posting Links in a Tweet URL: <https://support.twitter.com/articles/78124> [accessed 2016-10-13] [WebCite Cache ID 6lEkpAFm7]
44. Maas A. Learning word vectors for sentiment analysis. In: Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies. 2011 Presented at: 49th Annual Meeting of the Association for Computational Linguistics; 2011; Portland, Oregon.
45. Saif H, He Y, Alani H. Semantic sentiment analysis of Twitter. In: The Semantic Web ISWC 11th International Semantic Web Conference. 2012 Presented at: 11th International Semantic Web Conference; November 11-15, 2012; Boston, MA p. 11-524. [doi: [10.1007/978-3-642-35176-1\\_32](https://doi.org/10.1007/978-3-642-35176-1_32)]
46. Wang W, Chen L, Thirunarayan K, Sheth A. Harnessing Twitter "Big Data" for automatic emotion identification. 2012 Presented at: Privacy, Security, Risk and Trust (PASSAT), International Conference on Social Computing (SocialCom); 2012; Amsterdam, Netherlands. [doi: [10.1109/SocialCom-PASSAT.2012.119](https://doi.org/10.1109/SocialCom-PASSAT.2012.119)]
47. Li S, Xia W, Zong C, Huang C. A framework of feature selection methods for text categorization. 2009 Presented at: Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the Association for Computational Linguistics; August 2-7, 2009; Suntec, Singapore.
48. Pang B, Lee L, Vaithyanathan S. Thumbs up: sentiment classification using machine learning techniques. In: Proceedings of the ACL-2002 Conference on Empirical Methods in Natural Language Processing. 2002 Presented at: Association for Computational Linguistics; 2002; Stroudsburg, PA p. 79-86. [doi: [10.3115/1118693.1118704](https://doi.org/10.3115/1118693.1118704)]
49. Thelwall M, Buckley K, Paltoglou G, Cai D, Kappas A. Sentiment strength detection in short informal text. *J Am Soc Inf Sci* 2010;61:2544-2558. [doi: [10.1002/asi.21416](https://doi.org/10.1002/asi.21416)]
50. Hastie T, Tibshirani R, Friedman J. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. New York: Springer; 2009.

51. Sokolova M, Lapalme G. A systematic analysis of performance measures for classification tasks. *Information Processing & Management* 2009;45:427-437. [doi: [10.1016/j.ipm.2009.03.002](https://doi.org/10.1016/j.ipm.2009.03.002)]
52. Alvaro N, Conway M, Doan S, Lofi C, Overington J, Collier N. Crowdsourcing Twitter annotations to identify first-hand experiences of prescription drug use. *J Biomed Inform* 2015;58:280-287 [[FREE Full text](#)] [doi: [10.1016/j.jbi.2015.11.004](https://doi.org/10.1016/j.jbi.2015.11.004)] [Medline: [26556646](https://pubmed.ncbi.nlm.nih.gov/26556646/)]

## Abbreviations

**LR:** logistic regression

**NB:** naive bayes

**SVM:** support vector machines

**VADER:** Valence Aware Dictionary for sEntiment Reasoning

*Edited by T Sanchez; submitted 07.07.16; peer-reviewed by S Meaney, Z Zhang, N Alvaro; comments to author 26.07.16; revised version received 27.08.16; accepted 21.09.16; published 24.10.16.*

*Please cite as:*

*Daniulaityte R, Chen L, Lamy FR, Carlson RG, Thirunarayan K, Sheth A*

*“When ‘Bad’ is ‘Good’”: Identifying Personal Communication and Sentiment in Drug-Related Tweets*

*JMIR Public Health Surveill* 2016;2(2):e162

URL: <http://publichealth.jmir.org/2016/2/e162/>

doi: [10.2196/publichealth.6327](https://doi.org/10.2196/publichealth.6327)

PMID: [27777215](https://pubmed.ncbi.nlm.nih.gov/27777215/)

©Raminta Daniulaityte, Lu Chen, Francois R Lamy, Robert G Carlson, Krishnaprasad Thirunarayan, Amit Sheth. Originally published in *JMIR Public Health and Surveillance* (<http://publichealth.jmir.org>), 24.10.2016. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in *JMIR Public Health and Surveillance*, is properly cited. The complete bibliographic information, a link to the original publication on <http://publichealth.jmir.org>, as well as this copyright and license information must be included.



Original Paper

# Estimating the Duration of Public Concern After the Fukushima Dai-ichi Nuclear Power Station Accident From the Occurrence of Radiation Exposure-Related Terms on Twitter: A Retrospective Data Analysis

Naoki Nishimoto<sup>1\*</sup>, PhD; Mizuki Ota<sup>2\*</sup>, MSc; Ayako Yagahara<sup>3\*</sup>, PhD; Katsuhiko Ogasawara<sup>4</sup>, MBA, PhD

<sup>1</sup>Kagawa University Hospital, Clinical Research Support Center, Kita-gun, Japan

<sup>2</sup>Hokkaido University, Graduate School of Health Sciences, Sapporo, Japan

<sup>3</sup>Faculty of Health Sciences, Department of Radiological Technology, Hokkaido University of Science, Sapporo, Japan

<sup>4</sup>Hokkaido University, Faculty of Health Science School, Sapporo, Japan

\*these authors contributed equally

## Corresponding Author:

Katsuhiko Ogasawara, MBA, PhD

Hokkaido University

Faculty of Health Science School

Kita 12 zyou Nishi 5 chome

Kita-ku

Sapporo, 060-0812

Japan

Phone: 81 11 706 3409

Fax: 81 11 706 3409

Email: [oga@hs.hokudai.ac.jp](mailto:oga@hs.hokudai.ac.jp)

## Abstract

**Background:** After the Fukushima Dai-ichi Nuclear Power Station accident in Japan on March 11, 2011, a large number of comments, both positive and negative, were posted on social media.

**Objective:** The objective of this study was to clarify the characteristics of the trend in the number of tweets posted on Twitter, and to estimate how long public concern regarding the accident continued. We surveyed the attenuation period of the first term occurrence related to radiation exposure as a surrogate endpoint for the duration of concern.

**Methods:** We retrieved 18,891,284 tweets from Twitter data between March 11, 2011 and March 10, 2012, containing 143 variables in Japanese. We selected radiation, radioactive, Sievert (Sv), Becquerel (Bq), and gray (Gy) as keywords to estimate the attenuation period of public concern regarding radiation exposure. These data, formatted as comma-separated values, were transferred into a Statistical Analysis System (SAS) dataset for analysis, and survival analysis methodology was followed using the SAS LIFETEST procedure. This study was approved by the institutional review board of Hokkaido University and informed consent was waived.

**Results:** A Kaplan-Meier curve was used to show the rate of Twitter users posting a message after the accident that included one or more of the keywords. The term Sv occurred in tweets up to one year after the first tweet. Among the Twitter users studied, 75.32% (880,108/1,168,542) tweeted the word radioactive and 9.20% (107,522/1,168,542) tweeted the term Sv. The first reduction was observed within the first 7 days after March 11, 2011. The means and standard errors (SEs) of the duration from the first tweet on March 11, 2011 were 31.9 days (SE 0.096) for radioactive and 300.6 days (SE 0.181) for Sv. These keywords were still being used at the end of the study period. The mean attenuation period for radioactive was one month, and approximately one year for radiation and radiation units. The difference in mean duration between the keywords was attributed to the effect of mass media. Regularly posted messages, such as daily radiation dose reports, were relatively easy to detect from their time and formatted contents. The survival estimation indicated that public concern about the nuclear power plant accident remained after one year.

**Conclusions:** Although the simple plot of the number of tweets did not show clear results, we estimated the mean attenuation period as approximately one month for the keyword radioactive, and found that the keywords were still being used in posts at

the end of the study period. Further research is required to quantify the effect of other phrases in social media data. The results of this exploratory study should advance progress in influencing and quantifying the communication of risk.

(*JMIR Public Health Surveill* 2016;2(2):e168) doi:[10.2196/publichealth.5384](https://doi.org/10.2196/publichealth.5384)

## KEYWORDS

Twitter; social media; public concern; nuclear power plants; survival analysis; Kaplan-Meier estimate; infodemiology; radiation

## Introduction

### Spreading Concern About Radiation Exposure: Social Media and the Fukushima Dai-ichi Nuclear Power Station Accident in Japan

When the Fukushima nuclear power station accident occurred in Japan on March 11, 2011, concerns about radiation exposure arose. Continuous anxiety was assumed to affect people's physical and mental health, and the duration of fear or anxiety regarding the accident has remained unknown. An accurate estimation of the duration of concern would lead to recommendations related to public health action. To date, the duration of public concern has been difficult to measure because the data can be multi-scaled according to various topics and time scales. Many comments regarding the Fukushima accident have been posted on social media, such as Facebook and Twitter. It has become apparent that social media platforms encourage people to share their concerns regarding radiation exposure in their daily lives; a person's concern is highly correlated with the amount and length of a communication [1]. Thus, our primary interest was to quantify the amount of communication related to the specific topic of radiation exposure. Quantifying public anxiety plays an important role in public interventions [2].

Slovic surveyed the risk perception of lay people and experts concerning 30 activities, including nuclear power, and conducted factor analysis by plotting 81 hazards on a chart with the axes *dread risk* and *unknown risk* [3]. The concept *diagnostic X-rays* showed a low dread risk value, whereas *nuclear reactor accidents*, *nuclear weapons fallout*, and *radioactive waste* had high values for dread risk and medium values for unknown risk [3]. Slovic mentioned the Three Mile Island nuclear accident and concluded that the perception of risk was not related to the estimated social cost [3].

Mehta and Simpson-Housley reported a linear formula composed of factors such as trait-anxiety scores, gender, and whether people had children [4]. This research showed that expectation of a future nuclear power plant disaster was positively associated with high trait-anxiety scores, the female gender, and having children living in the household [4]. Until recently, it has not been possible to establish the effect of social media in generating public anxiety about nuclear power plant accidents and radiation exposure from radioactive waste. However, we have not been able to find any previous studies regarding the traits of terminology to capture written concerns about nuclear power plant accidents.

Exploring Twitter data, we aimed to generate a hypothesis regarding Twitter users' interest related to the Fukushima Dai-ichi Nuclear Power Station accident. Focusing on Twitter

users, we surveyed previous research on how users express their concerns regarding radiation exposure or a power station accident, and found only a small number of articles in the psychological literature [5]. With the emergence of social media, Twitter users have played a central role in health and the health care process, not only as recipients of health services, but also as initiators of positive personal health action. Users also bring the patient community together (eg, PatientsLikeMe.com, whose members manage home-based care for themselves and others) and act as citizens engaged in collaborative practices (ie, promoting proper sanitization and clean air) to ensure the health of their communities [6,7].

### Social Media in Health Care Research

Previous studies have explored evidence regarding the effects of social media on the perceptions of Twitter users. Approximately 90% of the Dutch population aged 12 years and older use the Internet, and 70% of these individuals are active on social media, particularly Facebook and Twitter (ie, the Web 2.0) [8]. According to the Japanese Ministry of Internal Affairs and Communications' *Communication Usage Trend Survey 2015*, Internet usage in Japan has shown an increasing trend, with approximately 82.2% of the population using the Internet and 62.6% using smartphones [9].

Twitter was introduced as an education support tool in the nursing domain for the first time in 2006 [10-13]. The neologism *infoveillance* (information + surveillance) was introduced by Eysenbach [14,15]. Novel methods for infoveillance are becoming available, such as mining, aggregating, and analyzing online textual data in real-time. Studies have shown that pandemic predictions or estimates of influenza-like illnesses derived from Twitter accurately track the reported levels of diseases [16-18]. Using Twitter data in conjunction with epidemiological data, Chew and Eysenbach reported on pandemic prediction of the H1N1 influenza outbreak, using double plot analyses to demonstrate that sharp increases in the absolute volume of H1N1-related tweets coincided with major H1N1 news events [18]. This finding demonstrated the usefulness of infodemiology techniques for pandemic prediction. In their 2014 study, Zhao et al reported a method for identifying influential users from Twitter data [19]. Several social media platforms are currently available, such as Twitter, Facebook, and LinkedIn. Due to its popularity, we focused on Twitter in this study.

### The Problem of Social Media-Based Big Data in Health Care

The concept of *big data* was introduced to the public by Douglas Laney, who discussed the problem of big data with respect to data volume, velocity, and variety [20]. At the time Laney wrote about this issue, big data had for the first time become

widespread in e-commerce; it subsequently infiltrated health care. The current definition of big data includes a further term: veracity [21]. Big data was regarded as a combination of large-scale, structured, and unstructured data, and the digital record of social infrastructure or social media was also regarded as part of this. The United States Federal Government has announced the challenges of big data and its mission for a big data program, which includes, “Informatics for Integrating Biology and the Bedside” developed by the National Library of Medicine, with the aim of creating more than 50 tools and approaches to facilitate the integration and exchange of informational byproducts of health care and biomedical research.

These tools were developed through open source sharing. Two major problems remain with this project. First, big data involves a large-scale volume of data, and powerful computing is required to handle this, yet professional medical staff, for example, may require real-time access to the data. Laney proposed the selection of data through sampling [20]. The American Statistical Association (ASA) published their policy entitled, “Discovery with data: leveraging statistics with computer science to transform science and society” to address these problems [21]. To handle large-volume data, the ASA reviewed techniques and started with data visualization. The statistics community has a long history of developing data visualization techniques, not just histograms and scatterplots, but also techniques such as trellis plots and dynamic graphs. The ASA has also introduced modern visualization techniques such as treemaps and other techniques for visualizing network data; these approaches are going to be heavily in demand, and new ways of visualizing complex data with specific properties will need to be developed [22]. Causal inference is also covered in the policy of the ASA, and summarizing data using graphs plays an important role in estimating causal inference. Techniques used in data mining do not appear to be synthesized with conventional statistical methodology; the ASA has proposed combining data mining techniques with visualization, and have claimed that this approach has the potential to exceed the power of either field alone.

Second, it is difficult to identify the individual who posted a comment on social media, and existing statistical methodologies can be difficult to use, as it may not be clear whether the observation was independent or not. The problem of defining a *study population* remains in social media research, as some people post comments using a user identification (ID) with multiple users, or a user ID from an automated tweeting program. Twitter has a growing user population and is open in nature, which has made it an ideal target for exploitation by automated tweeting programs, known as *bots*. Like bots in other Web applications such as Internet chatrooms, blogs, and online games, bots have become common on the Twitter platform [23-25].

## Motivation

We have focused on public concern related to the risk of a nuclear power plant accident and concern about the negative effects of radiation on the human body. Public concern is affected by the media (eg, TV, radio, or Internet), making this study unique from an informatics perspective. Over the course

of this study we observed a widespread concern that has not been quantified. Some phrases related to radiation exposure (ie, *air dose rate* or *no radiation particle spread*) that were mentioned in microblogs may have been used to infer the status and safety of specific conditions in geographically distinct regions. However, it is not well established when these words arise or how long public concern lasts among Twitter users.

## Purpose

The duration of public concern has thus far been difficult to measure, because the data can be multi-scaled according to various topics and time scales. Exploring Twitter data required making assumptions regarding Twitter users' interest in the Fukushima Dai-ichi Nuclear Power Station accident.

The purpose of this study was to clarify the characteristics of the trend in the number of tweets posted on Twitter, and to estimate how long public concern about the accident lasted. We surveyed the attenuation period of radiation exposure-related terms for the first time as a surrogate endpoint for the duration of concern.

## Methods

### Twitter Data and Tools

We retrieved Japanese Twitter data from March 11, 2011 to March 10, 2012 as social networking service data from Twitter, Incorporated [26]. Twitter was created in March 2006, and the service rapidly gained worldwide popularity; by 2012, more than 100 million users were posting 340 million tweets per day [27]. Each tweet contains a Tweet ID, a textual part (restricted to 140 characters), and the tweet's date and time. We obtained 18,891,284 tweets and 143 variables, including date and time of tweet, tweet text, user ID, and user names. The file size totaled 17.2 gigabytes. The data was in comma-separated values (CSV) format.

We also used the programming language Java (jre1.8.0\_25) with Eclipse Luna 4.4.0, an integrated development environment, for data handling. The statistical packages SAS 9.4 and JMP 11 pro (both SAS Institute Inc., Cary, NC) were used for estimating the survival curve. We used the HP Z420 workstation (Hewlett-Packard Company) with 3.7 gigahertz central processing unit and 64 gigabyte memory.

Figure 1 depicts a flowchart for the construction of the dataset. We obtained the Twitter data in CSV format and divided the information into 10 subsets using an original Java program, because it was easy to detect errors when converted into the SAS datasets. To handle the Twitter data using the SAS, we formatted the variable names as alphanumeric. We used the frequency (FREQ) procedure (an SAS analysis program) for count and categorical data, and obtained time series data for the tweet count per day.

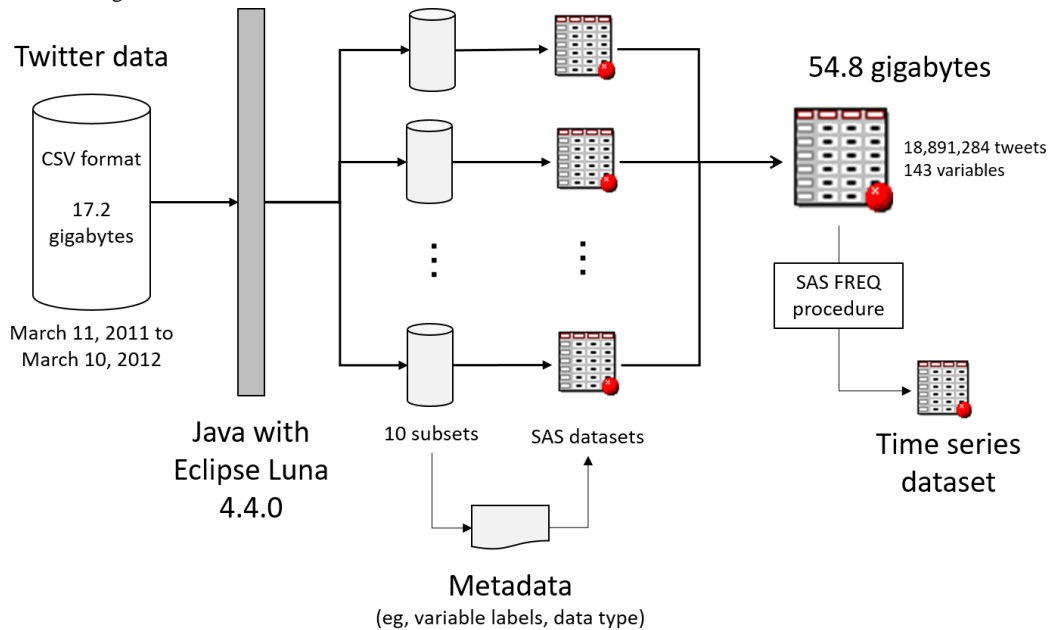
We intended to quantify public concern regarding radiation exposure from the Fukushima Dai-ichi Nuclear Power Station accident. Although we treated the tweet count as a measure of public concern, there appeared to be a gap between the concerns of Twitter users and the tweet count. Consequently, we showed the correlation by plotting air dose rate data against tweet count

as evidence of surrogacy. Data for the air dose rate around the Fukushima Dai-ichi Nuclear Power Station were available to the public on the Tokyo Electric Power Company's webpage [28], and as of June 2015 these data had been updated only in Japanese. The Tokyo Electric Power Company established eight

monitoring posts (MPs) to estimate the air dose rate within the area of the power station.

This study was approved by the institutional review board of Hokkaido University, Health Science School. Informed consent was waived because all records were anonymized and we surveyed data retrospectively.

**Figure 1.** Flowchart showing data construction.

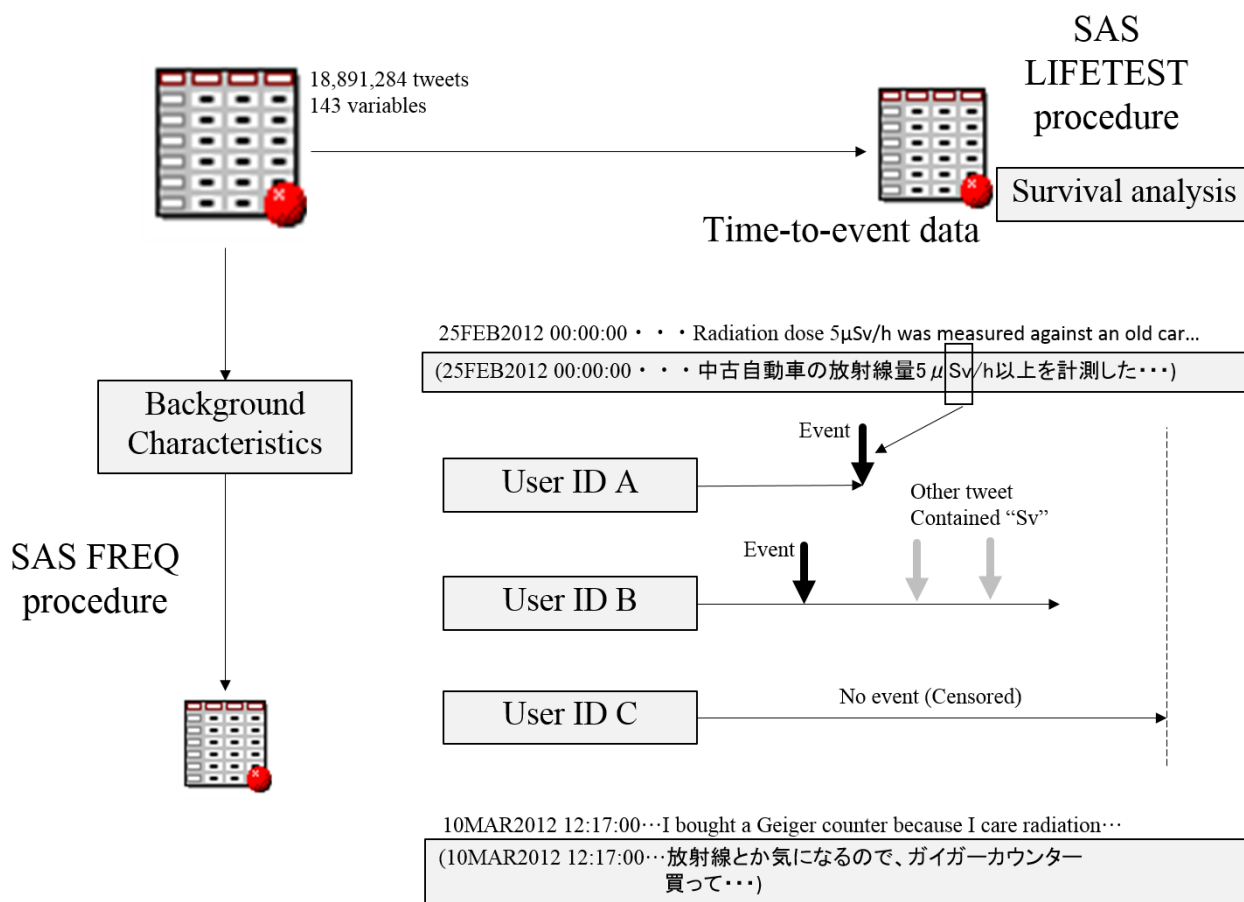


### Selection of Keywords Related to Concern About Radiation Exposure

Keyword selection played an essential role in this research. It was important to select keywords that were sensitive in detecting the concern about radiation risk. Figure 2 shows event detection using the dictionary. When the mass media broadcast information related to radiation exposure, the words *radiation* or *radioactive* were used frequently, along with the radiation units *Sievert (Sv)*, *gray (Gy)*, and *Becquerel (Bq)*. We selected *radiation*, *radioactive*, *Sv*, *Bq*, and *Gy* as keywords to estimate the attenuation period. The keywords *radiation* and *radioactive* were in Japanese and the radiation units were used in their normal format. Natural language processing techniques are frequently used for text analyses in radiology, or to identify patient smoking status; however, it was difficult to merge the

two different words into a single concept in the Twitter datasets because the free text part of Twitter data was restricted to 140 characters, leaving little contextual information available to use in this case [29,30]. Thus, we maintained the reproducibility based on the keyword selection using multiple concept names. Using the keywords, we matched the concept names with those terms in the free text portion of the Twitter datasets. The mean proportions of tweets that contained the keywords *radiation* or *radioactive* were plotted on an hourly basis.

When a Twitter user posted a tweet containing one of the keywords for the first time, we regarded this as an *event*, and analyzed it using the SAS LIFETEST procedure. Background characteristics such as mean count per day were plotted as a time series plot using the SAS FREQ procedure. This analysis would normally show a great deal of censored data, so we suppressed the censored plot on the Kaplan-Meier curve.

**Figure 2.** Flowchart for the identification of the primary endpoint.

### Estimating Public Concern Using Survival Analysis

The survival analysis technique is useful for handling and evaluating time series data. In the field of biostatistics, an event in this analysis indicates the death of a subject, and survival time refers to a period with no event (and is thus interpreted as, for example, disease-free survival or progression-free survival). Many textbooks related to survival analyses have been published [31,32], but the use of this methodology in the present study was quite different from its use in biostatistics.

When we conducted the survival analysis, we could not determine the starting time point for the observation because the individual who tweeted the event could not be followed by the authors in social media; this issue highlights one important difference between big data analysis and other types of research, such as clinical trials. We assumed that observation started on March 11, 2011 in a pseudo manner. Although some ambiguity in Twitter data is inevitable, the volume of data can reduce this effect [21]. We assumed the onset of a tweet and defined an event as when the  $k^{\text{th}}$  keyword occurred in the free text part of a tweet for the first time from a given user ID. Thus, the survival function was computed as follows:

Here,  $i$  indicates the  $i^{\text{th}}$  day after March 11, 2011. Date and time were saved as a continuous variable in SAS. We estimated the duration of concern in seconds. The duration of concern could be attenuated with time; we estimated the mean attenuation period with  $S(t_{ik})$ .

### Results

#### Twitter Data Characteristics

Table 1 shows the characteristics of tweets and Twitter IDs from March 11, 2011 to March 10, 2012. The median tweet count was 43,345 tweets/day, ranging from 25,769 to 388,984 tweets/day. The histogram was skewed right with a skewness of 5.85. The data included 1,168,543 user IDs (16.2 tweets/user ID). Available usernames numbered 879,210, but there were 125,363 usernames that were missing data. We calculated the mean number of IDs per day as 23,626.23, and 2.19 tweets per ID were posted per day. The tweet count per day showed a variation just after March 11, 2011, with spikes in September and October of 2011 in the time series plot (Figure 3). From a total of 18,891,284 tweets, 9,673,756 (51.21%) were retweets of original messages.



**Table 1.** Characteristics of tweets and Twitter IDs on Twitter from March 11, 2011 to March 10, 2012.

	Tweets	Twitter IDs
Total	18,891,284	1,168,543
Mean/day, n (95% CI)	51,616 (48,044-55,187)	23,626 (21,695-25,557)
Standard deviation	34,747	18,786
Skewness	5.85	5.04
Kurtosis	43.15	32.07
Minimum/day, n	25,769	9421
25 <sup>th</sup> percentile/day, n	38,766	15,621
Median/day, n	43,345	18,994
75 <sup>th</sup> percentile/day, n	51,703	23,448
Maximum/day, n	388,984	187,291

The air dose rate showed some variation depending on the wind direction or climatic condition. The MPs output the data for the air dose rate at 10 minute intervals. We downloaded the publicly available data and selected the eight MPs to show air dose rate. [Figure 3](#) shows the mean hourly frequency of tweets plotted for each day of the week. The frequency of tweeting decreased from 4 o'clock to 5 o'clock a.m., and then peaked at 12 o'clock noon. There was a further increase at 8 o'clock p.m. We observed that the overall pattern each day was almost constant, regardless of the day of the week.

[Figure 4](#) and [Table 2](#) show the similarity in the daily tweet count and the mean MP air dose rate. [Figure 4](#) displays a time series plot of the normalized tweet count over the entire period from March 11, 2011 to March 10, 2012, and the air dose radiation

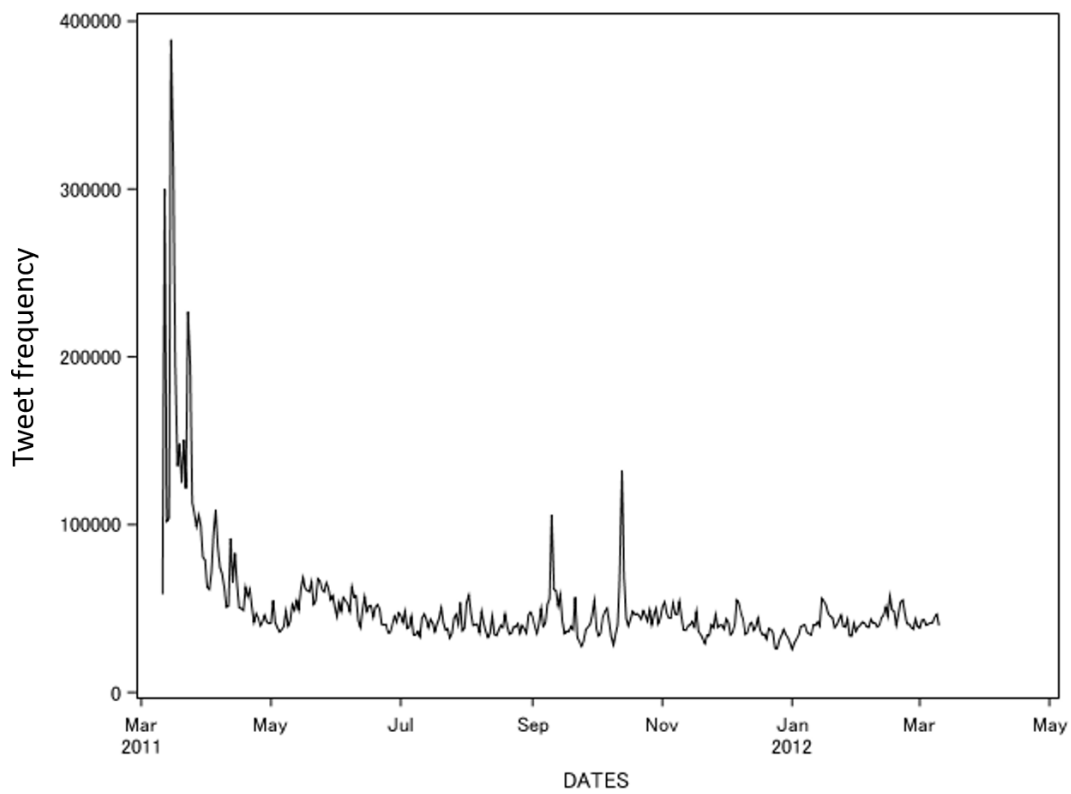
rate measured within the Fukushima Dai-ichi Nuclear Power Station. We observed a peak shortly after March 11, 2011 and relatively small peaks in September and October of 2011; there was also some daily variation. The air dose values observed at an MP were averaged per day, based on the data obtained at 10 minute intervals. Similarity was observed between the air dose rate and the tweet count.

[Figure 5](#) shows the plot of the keyword, including proportion averaged by month for each day of the week. The proportion plot including *radiation* showed some positive and negative peaks at approximately 7 o'clock a.m. and 9 o'clock p.m. ([Figure 5 A](#)); the plot of *radioactive* appeared to be the opposite of *radiation* ([Figure 5 B](#)).

**Table 2.** Sum of squared differences between the normalized MPs and normalized tweet counts. The normalized air dose rate of MP-5 and MP-6 showed a good fit to the normalized tweet count plot.

Monitoring post	Sum of squared difference
MP-1	28.72
MP-2	62.51
MP-3	14.35
MP-4	11.59
MP-5	4.66
MP-6	8.24
MP-7	20.93
MP-8	25.40

**Figure 3.** Time series plot of the tweet count over the period from March 11, 2011 to March 10, 2012. The frequency of tweeting peaked at 388,984 per day.



**Figure 4.** Time series plot of the normalized tweet count from March 11, 2011 to March 10, 2012.

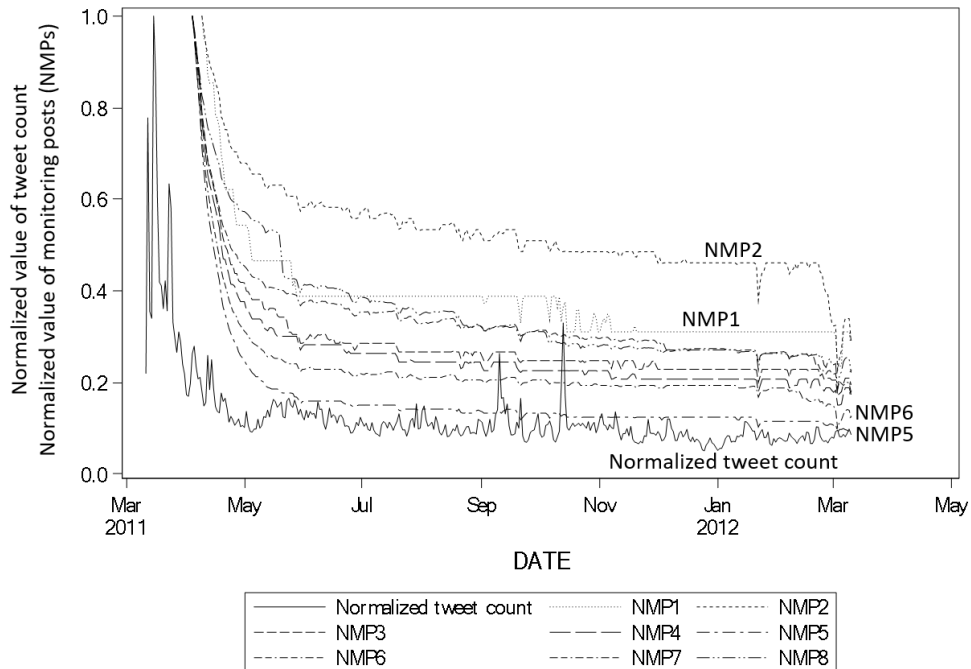
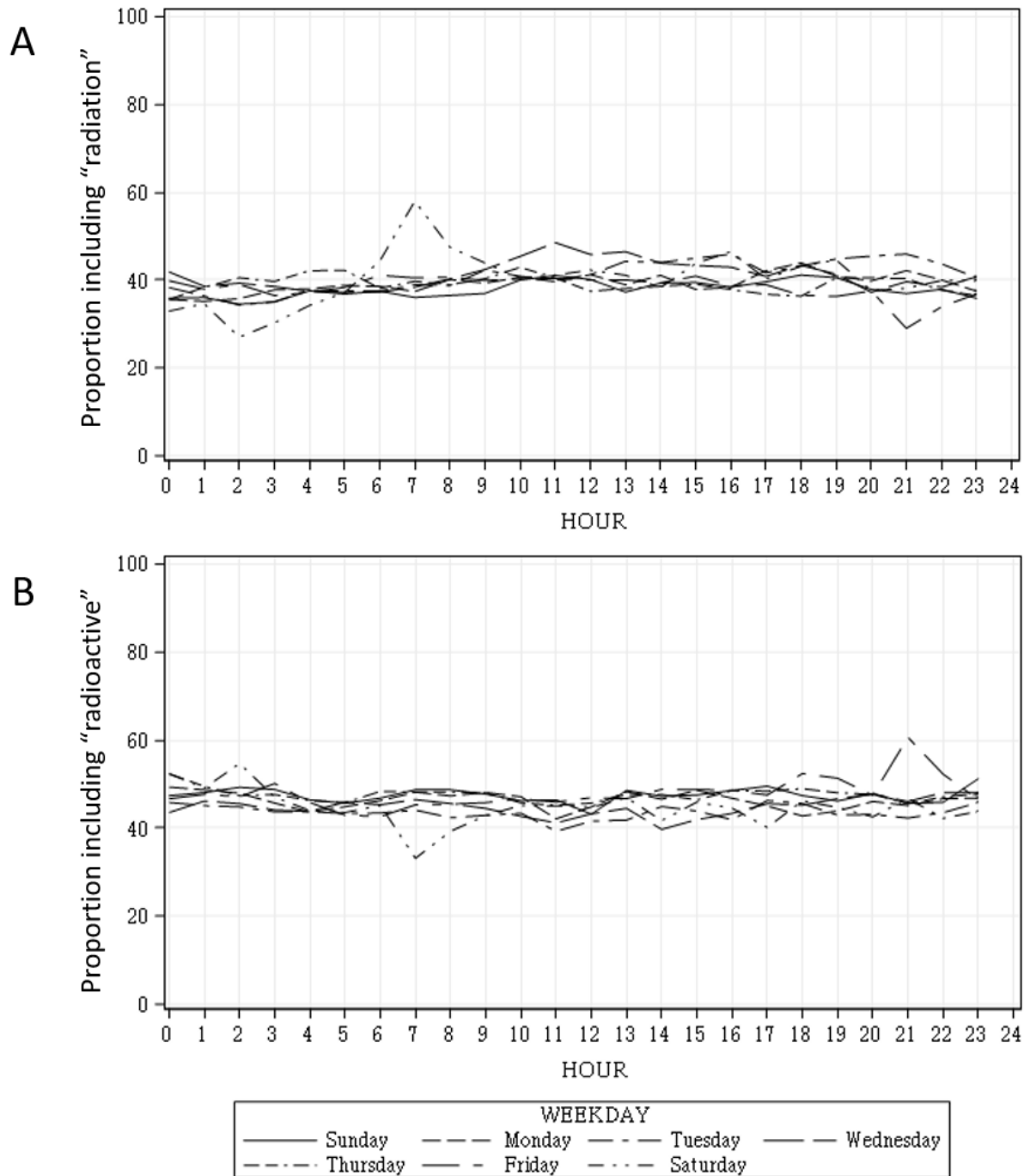


Figure 5. Plot of the keywords radiation and radioactive, including proportion averaged by month for each day of the week.



**Survival Analysis of Twitter Events: Estimating the Attenuation Period for the Keywords**

We conducted survival analyses for tweets that contained the keywords. Figure 6 shows the Kaplan-Meier curve of the first tweet rate for tweets that included the keywords radiation or radioactive. The curve for radioactive dropped below that of radiation. People continued to tweet the word radioactive a

year after the first tweet, and of the 1,168,542 user IDs, 880,108 (75.32%) tweeted radioactive (Figure 6). The mean attenuation periods were 63.4 days (standard error [SE] 0.152) and 31.9 days (SE 0.096) for radiation and radioactive, respectively (Table 3). The mean attenuation period for these was much shorter than those for the radiation unit keywords. The number of events for Sv was twice the number for Bq and five times that of Gy (Table 3). We treated the remaining data as censored.

**Table 3.** Number of events and estimated attenuation period derived from the Twitter data.

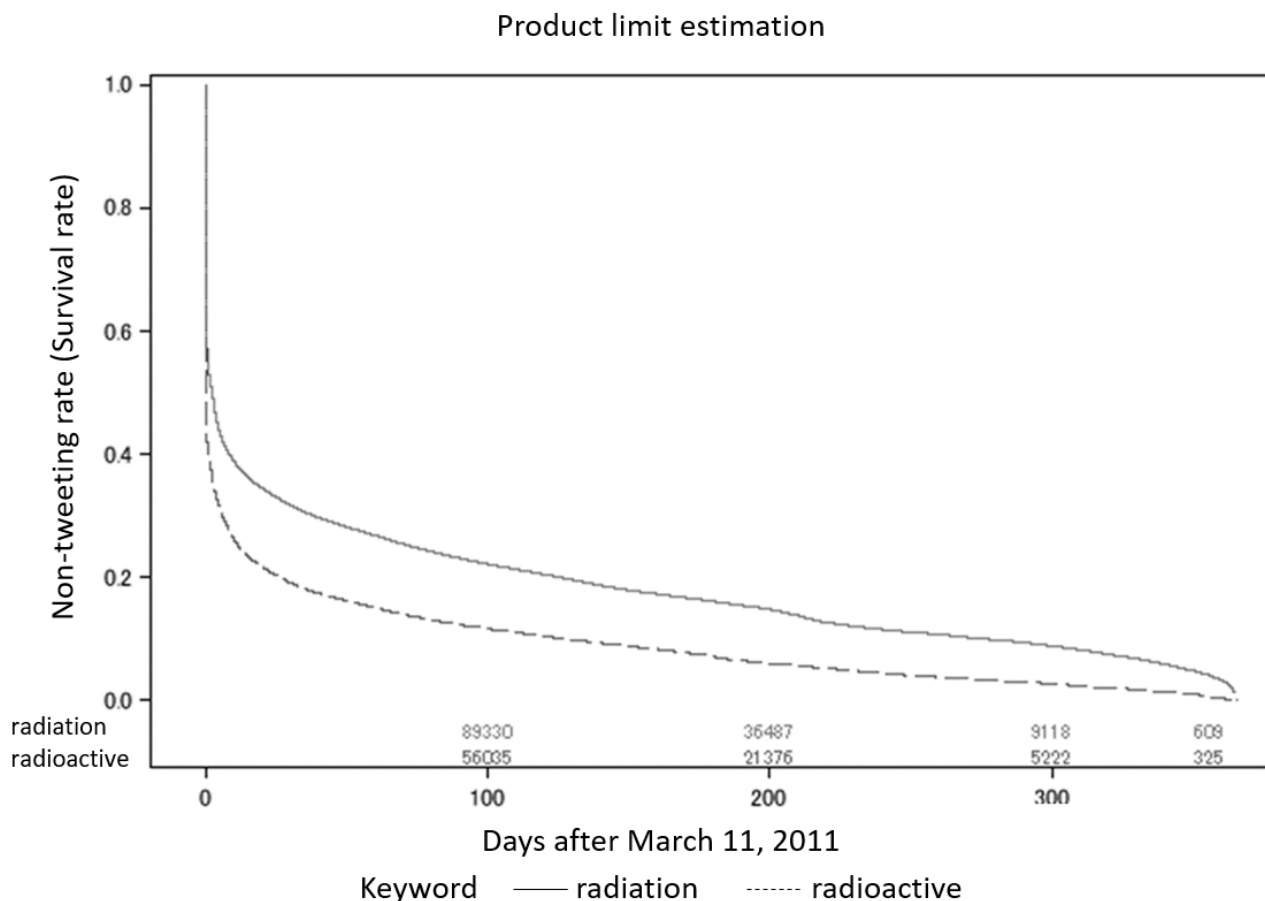
Keyword (Japanese)	Twitter ID count	Event	Censor	% Censored	Mean, days	95% Lower Control Limit	95% Upper Control Limit
radioactive (放射能)	1,168,542	880,108	288,434	24.68	31.9	31.7	32.1
radiation (放射線)	1,168,542	710,924	457,618	39.16	63.4	63.1	63.7
Sv	1,168,542	107,522	1,061,020	90.8	300.6	300.2	301.0
Bq	1,168,542	53,034	1,115,508	95.46	330.3	330.0	330.6
Gy	1,168,542	17,111	1,151,431	98.54	354.6	354.4	354.8

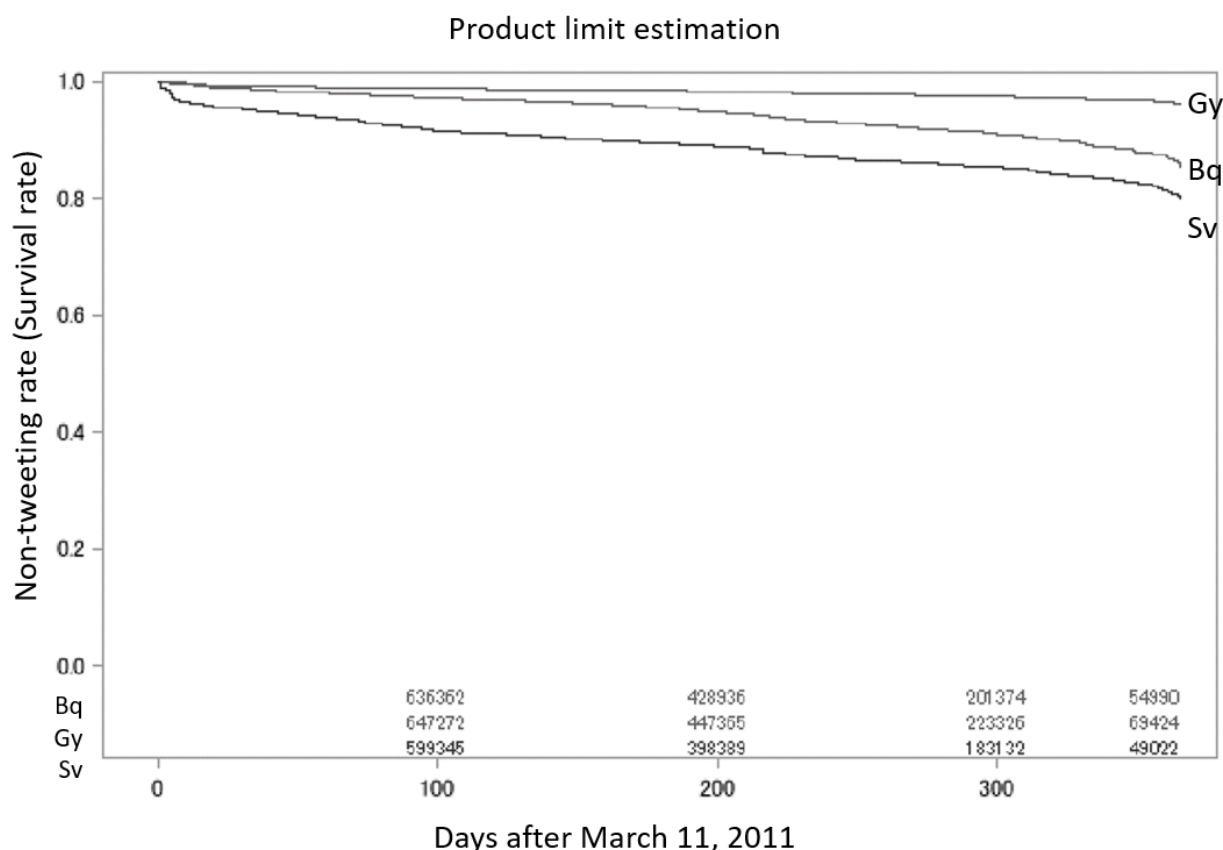
Figure 7 shows the Kaplan-Meier curve of the first tweet rates for tweets that contained the keywords Sv, Gy, and Bq. Users continued to tweet messages that included Sv up to one year after the first tweet. Of the 1,168,542 user IDs, 107,522 (9.20%) tweeted Sv. We were not able to obtain the median for the first tweet rate between March 11, 2011 and March 10, 2012 because the event rate decreased to approximately 0.8 within one year.

We observed a steep decrease in nontweeting rate during the first 7 days after March 11, 2011 (Figure 6). This finding

indicated the occurrence of automatic tweeting from bots. Tweets by bots can be identified from their contents; however, no specific data handling was carried out for these user IDs because natural language processing is arduous in this context, and we decided that bot identification was beyond the of scope in this study. There was a gradual decrease in nontweeting rate for tweets that referenced radiation units, and no sudden changes were identified during the observation period in (Figure 7). Table 3 shows the counts for the events and censored data.

**Figure 6.** Kaplan-Meier curves of the tweet rates for tweets including the keywords radiation or radioactive.



**Figure 7.** Kaplan-Meier curves of the tweet rates of tweets including the keywords Bq, Gy, or Sv.

## Discussion

### Tweet and Twitter ID Trends Became Stable in the Count Plots

The maximum number of Twitter IDs tweeting a keyword in a single day was approximately 187,000. At its minimum, the number of Twitter IDs was 9,421 per day (approximately 5% of the maximum), while the median was 18,994 per day (approximately 10% of the maximum). However, the 75<sup>th</sup> percentile (23,448 IDs/day) was not greatly different from the

median (Table 1). We observed that the maximum number of IDs per day was an outlier.

Although the count of Twitter IDs did not directly indicate the number of users, only 1.00% of the Twitter users appeared to be bots (Table 4). The effect of bots could be estimated, and we decided that no specific data handling would be given to tweets from suspected bots. We could not establish the Twitter share in social media among the population; however, it has been estimated that 79.1% of the population of Japan were Internet users in 2011 [33]. We can therefore conclude that the survey from Twitter showed a trend pertaining to the entire population.

**Table 4.** Percentiles of tweet counts per Twitter ID (N=1,168,542).

Level	Percentile	Number (n)
Maximum	100%	71,793
	99%	243
	95%	42
	90%	18
3 <sup>rd</sup> quartile	75%	6
Median	50%	2
1 <sup>st</sup> quartile	25%	1
	10%	1
	5%	1
	1%	1
Minimum	0%	1



## Time Series Plot and Combination with Other Epidemiological Data

The time series plot of tweets (Figure 3) showed a surge just after March 11, 2011, and then a decrease until May 2011, after which time the number of tweets appeared to be stable. We observed some daily variation, and two peaks from September to November of 2011. We assumed that the number of tweets tended to follow the decrease in the radiation dose (Figure 4); however, it was difficult to estimate the effects of mass media. We assumed that public concern was burned out immediately, despite the fact that the risk of radioactive exposure persisted. The Tweet counts may have decreased because Twitter users did not live near the Fukushima Dai-ichi site, but further study would be required to determine the location of the users.

When we examined the event related to the surge of tweets on September 10, 2011 (Figure 3), we discovered that the Japanese Minister of Economy, Trade, and Industry made a negative comment about the Fukushima Dai-ichi Nuclear Power Station accident in an interview, and the Japanese mass media broadcast this statement. Contents of Twitter data at that time indicated that many tweets mentioned or retweeted this news story. On October 13, 2011, we found that the tweet count increased because a sealed radioactive source was found under a house in Tokyo, and the mass media broadcast this news. However, this incident was not directly related to the Fukushima Dai-ichi Nuclear Power Station accident, although Twitter users still interpreted it as news about radiation exposure.

The time series plot in Figure 3 indicated that the tweeting trend was somewhat ambiguous as it related to how many users were actually concerned about radiation exposure. From another point of view, we observed that a simple plot of the tweet count indicated that public concern was high at first, and then decreased and became stable.

## Estimating the Attenuation Period

We observed that the first tweet counts differed in number, with *Sv* being mentioned the most, followed by *Bq* and *Gy*. These counts indicated that Twitter users observed mass media using the radiation unit *Sv*, which is related to the effects of radiation exposure on humans. Approximately 1.00% of Twitter user IDs posted a large number of tweets related to our keywords (more than 200 tweets over the year); we checked the contents of these messages and determined that they had the traits of auto-posted messages, such as daily radiation reports. To estimate the activity from these indexes, the problem of nonhuman user IDs remained. Some researchers have reported using natural language processing techniques based on the contents of tweets; however, it remained difficult to identify nonhuman users [24,25].

There appeared to be another reason for the difference between the mean and median: some IDs posted no message during the observed period. The mean was therefore assumed to be considerably reduced, but a sensitivity analysis would be required to clarify this issue. To predict the changes in a trend over a period of time, cyclic variation reduction methodology is required from another domain, such as time series analyses.

We could not estimate the median attenuation period from the Kaplan-Meier plot. However, we estimated the nonfirst tweeting rate (survival rate) at 365 days after March 11, 2011. A tweet (containing the keyword) that was the first for that user ID was observed, and there was a reduction in the nonterm tweeting rate within the first 7 days. The reduction at this time point was highly dependent on the users that we identified as bots. We did not exclude these in our research because this led to an overestimation bias of the nontweeting rate. Figure 6 and Figure 7 show that the nontweeting rate decreased gradually with time. In reference to the time series plot in Figure 3, only a small number of tweets were shown in the plot, and the trend showed a steep peak within the first 7 days, followed by a stable trend. When we focused on the Kaplan-Meier plots (Figure 6 and Figure 7), we clearly observed that the tweet that mentioned keywords for the first time continued to be posted for a year.

Figure 6 and Figure 7 show that the rate of user IDs that posted a keyword increased one year after the Fukushima Dai-ichi Nuclear Power Station accident. The simple time series plot in Figure 3 showed the peak immediately after the accident and the subsequent decrease in tweet count, demonstrating that public concern became stable. However, Twitter users had fears or concerns about nuclear exposure, as seen in the event probability plot (Figure 6 and Figure 7). When we drew a time series plot of the tweet count, the information related to public concern was hidden. We must consider that not every tweet expressed a negative attitude towards radiation exposure, although public concern could be a denial of radiation exposure. Nuclear power plant activity ceased throughout Japan after the accident at Fukushima, and we believe this fact supports tweet contents that reflected a negative attitude towards nuclear energy. Thus, using the event probability plot by time series was a powerful technique for this purpose.

## Limitations

We used Twitter data to detect public concern about radiation exposure; however, some ambiguity was shown in our study. If we had been following the techniques of surveillance in an epidemiological study, we should have identified each participant in the study and followed them with a time series. We had Twitter user IDs, and these could be shared or checked by an automated program. We assumed that this fact is not accepted in the epidemiological domain, although we would like to emphasize the effectiveness of this method for early stage prediction and the quantification of public concern about events in society.

In addition, we tracked the radiation-related keywords, but we did not consider the context in which they were used. The positivity or negativity of the tweets should be added into the analysis, which would allow for greater precision when analyzing the data.

## Future Work

We selected very few keywords for our analyses, and the estimation of Twitter users' interest had some limitations. Natural language processing could be used for content analysis as an engineering task. Similarly, direct emotions (eg, concern or fear) could also be challenging in the health care domain.

The effect of low dose radiation exposure was not clear in the present study. We would like to quantify these emotions as an effect, in addition to the biological effects on human health. Using survival analysis is still problematic because the definition of onset was not clear, and this may be difficult to accept as an observational study in the epidemiological field. Further study is required to adapt the current methodology.

## Conclusions

Although the simple plot of the tweet count did not show clear results, we estimated the mean attenuation period to be approximately one month for the keyword *radioactive*, and found that this keyword was still being used at the end of the investigation period. Further research is required to quantify the effects of other phrases in social media data. The results of this exploratory study should advance progress in influencing and quantifying the communication of risk.

## Acknowledgments

The authors are grateful to the anonymous referees for their insightful comments and feedback. This work was partially supported by Grant-in-Aid for Challenging Exploratory Research, 26670293.

## Conflicts of Interest

None declared.

## References

1. Jungherr A. Analyzing political communication with digital trace data: the role of Twitter messages in social science research. Heidelberg, Germany: Springer; Jun 25, 2015.
2. Suzuki T. The Problems of Analyzing the Factors behind People's Anxiety Based upon the Use of True-False Test in Regards to Their Knowledge of Radiation. JJSC 2014 Jun;15:3-16 [article in Japanese] [FREE Full text]
3. Slovic P. Perception of risk. Science 1987 Apr 17;236(4799):280-285. [Medline: 3563507]
4. Mehta MD, Simpson-Housley P. Trait anxiety and perception of a potential nuclear power plant disaster. Psychol Rep 1994 Feb;74(1):291-295. [doi: 10.2466/pr0.1994.74.1.291] [Medline: 8153221]
5. Chang BL, Friedman CP, Kukafka R, Eysenbach G, Harrison JH, Lewis D, et al. In: Lewis D, Eysenbach G, Kukafka R, Stavri PZ, Jimison H, editors. Consumer health informatics: informing consumers and improving health care. New York: Springer; 2005.
6. Frost J, Okun S, Vaughan T, Heywood J, Wicks P. Patient-reported outcomes as a source of evidence in off-label prescribing: analysis of data from PatientsLikeMe. J Med Internet Res 2011 Jan 21;13(1):e6 [FREE Full text] [doi: 10.2196/jmir.1643] [Medline: 21252034]
7. Schroeder EB, Desai J, Schmittiel JA, Paolino AR, Schneider JL, Goodrich GK, et al. An innovative approach to informing research: gathering perspectives on diabetes care challenges from an online patient community. Interact J Med Res 2015 Jun 30;4(2):e13 [FREE Full text] [doi: 10.2196/ijmr.3856] [Medline: 26126421]
8. Centraal Bureau voor de Statistiek. Seven in ten internet users active on social media. 2013 Apr 10. URL: <http://www.cbs.nl/en-GB/menu/themas/vrije-tijd-cultuur/publicaties/artikelen/archief/2013/2013-3907-wm.htm> [accessed 2016-11-07] [WebCite Cache ID 6lphjhyAA]
9. Ministry of Internal Affairs and Communications. Information & Communications Statistics Database. 2013. URL: <http://www.soumu.go.jp/johotsusintokei/statistics/statistics05b1.html> [accessed 2016-11-07] [WebCite Cache ID 6lphu08Ex]
10. Skiba DJ. Nursing education 2.0: Twitter & tweets. Can you post a nugget of knowledge in 140 characters or less? Nurs Educ Perspect 2008;29(2):110-112. [Medline: 18459627]
11. Thames G. Twitter as an educational tool. J Child Adolesc Psychiatr Nurs 2009 Nov;22(4):235. [doi: 10.1111/j.1744-6171.2009.00208.x] [Medline: 19930305]
12. Bristol TJ. Twitter: consider the possibilities for continuing nursing education. J Contin Educ Nurs 2010 May;41(5):199-200. [doi: 10.3928/00220124-20100423-09] [Medline: 20481418]
13. Mistry V. Critical care training: using Twitter as a teaching tool. Br J Nurs 2011;20(20):1292-1296. [doi: 10.12968/bjon.2011.20.20.1292] [Medline: 22068003]
14. Eysenbach G. Infodemiology: tracking flu-related searches on the web for syndromic surveillance. AMIA Annu Symp Proc 2006:244-248 [FREE Full text] [Medline: 17238340]
15. Eysenbach G. Infodemiology and infoveillance: framework for an emerging set of public health informatics methods to analyze search, communication and publication behavior on the Internet. J Med Internet Res 2009 Mar 27;11(1):e11 [FREE Full text] [doi: 10.2196/jmir.1157] [Medline: 19329408]
16. St Louis C, Zorlu G. Can Twitter predict disease outbreaks? BMJ 2012 May 17;344:e2353. [Medline: 22597352]
17. Gesualdo F, Stilo G, Agricola E, Gonfiantini MV, Pandolfi E, Velardi P, et al. Influenza-like illness surveillance on Twitter through automated learning of naïve language. PLoS One 2013;8(12):e82489 [FREE Full text] [doi: 10.1371/journal.pone.0082489] [Medline: 24324799]

18. Chew C, Eysenbach G. Pandemics in the age of Twitter: content analysis of Tweets during the 2009 H1N1 outbreak. *PLoS One* 2010 Nov 29;5(11):e14118 [FREE Full text] [doi: [10.1371/journal.pone.0014118](https://doi.org/10.1371/journal.pone.0014118)] [Medline: [21124761](https://pubmed.ncbi.nlm.nih.gov/21124761/)]
19. Zhao K, Yen J, Greer G, Qiu B, Mitra P, Portier K. Finding influential users of online health communities: a new metric based on sentiment influence. *J Am Med Inform Assoc* 2014 Oct;21(e2):e212-e218 [FREE Full text] [doi: [10.1136/amiajnl-2013-002282](https://doi.org/10.1136/amiajnl-2013-002282)] [Medline: [24449805](https://pubmed.ncbi.nlm.nih.gov/24449805/)]
20. Laney D. 3D Data Management: Controlling Data Volume, Velocity, and Variety. META Group. 2001. URL: <http://blogs.gartner.com/doug-laney/files/2012/01/ad949-3D-Data-Management-Controlling-Data-Volume-Velocity-and-Variety.pdf> [accessed 2016-11-07] [WebCite Cache ID 6lq9vvdvp]
21. Rudin C, Dunson D, Irizarry R, Ji H, Laber E, Leek J, et al. Discovery with Data: Leveraging Statistics with Computer Science to Transform Science and Society. 2014. URL: <http://www.amstat.org/ASA/Science-Policy-and-Advocacy/home.aspx> [accessed 2016-11-07] [WebCite Cache ID 6lqAXvYaD]
22. Shneiderman B, Plaisant C. Treemaps for space-constrained visualization of hierarchies, Including the History of Treemap Research at the University of Maryland. 2014. URL: <http://www.cs.umd.edu/hcil/treemap-history/> [accessed 2016-11-16] [WebCite Cache ID 6m4jgVTRy]
23. Gianvecchio S, Wu Z, Xie M, Wang H. Battle of Botcraft: fighting bots in online games with human observational proofs. In: Proceedings of the 16th ACM conference on Computer and communications security. 2009 Nov 09 Presented at: The 16th ACM conference on Computer and communications security, ACM: Chicago, Illinois, USA; November 09 - 13 p. 256-268.
24. Gianvecchio S, Xie M, Wu Z, Wang H. Measurement and classification of humans and bots in internet chat. In: Proceedings of the 17th conference on Security symposium. 2008 Jul 28 Presented at: The 17th conference on Security symposium. USENIX Association; San Jose, CA; July 28-August 1 p. 155-169.
25. Stone-Gross B, Cova M, Cavallaro L, Gilbert B, Szydlowski M, Kemmerer R. Your botnet is my botnet: analysis of a botnet takeover. In: Proceedings of the 16th ACM conference on Computer and communications security. 1653738: ACM; 2009 Nov 09 Presented at: The 16th ACM conference on Computer and communications security; November 09 - 13; Chicago, Illinois, USA p. 635-647.
26. Twitter Inc. Twitter Inc. 2016. URL: <https://twitter.com> [accessed 2016-11-16] [WebCite Cache ID 6m4kGm6iL]
27. Novak A, El-Burki IJ. Defining Identity and the Changing Scope of Culture in the Digital Age. Hershey, PA: IGI Global; 2016.
28. Tokyo Electric Power Company. Mesuring status of monitoring post inside Fukushima daiichi nuclear power station. 2016. URL: <http://www.tepco.co.jp/nu/fukushima-np/f1/index-j.html> [accessed 2016-11-07] [WebCite Cache ID 6lqFJ8sXT]
29. Uzuner O, Goldstein I, Luo Y, Kohane I. Identifying patient smoking status from medical discharge records. *J Am Med Inform Assoc* 2008;15(1):14-24 [FREE Full text] [doi: [10.1197/jamia.M2408](https://doi.org/10.1197/jamia.M2408)] [Medline: [17947624](https://pubmed.ncbi.nlm.nih.gov/17947624/)]
30. Carrell D, Miglioretti D, Smith-Bindman R. Coding free text radiology reports using the Cancer Text Information Extraction System (caTIES). *AMIA Annu Symp Proc* 2007 Oct 11:889. [Medline: [18693990](https://pubmed.ncbi.nlm.nih.gov/18693990/)]
31. Kalbfleisch J, Prentice R. The statistical analysis of failure time data. Hoboken, NJ: Wiley; 2002.
32. Armitage P, Berry G, Matthews J. Statistical methods in medical research. Malden, MA: Blackwell Science; 2002.
33. Ministry of Internal Affairs and Communications. Communications Usage Trend Survey in 2011. 2012. URL: [http://www.soumu.go.jp/johotsusintokei/tsusin\\_riyou/data/eng\\_tsusin\\_riyou02\\_2011.pdf](http://www.soumu.go.jp/johotsusintokei/tsusin_riyou/data/eng_tsusin_riyou02_2011.pdf) [accessed 2016-11-17] [WebCite Cache ID 6m6MxkEdA]

---

## Abbreviations

**ASA:** American Statistical Association  
**Bq:** Becquerel  
**CSV:** comma-separated values  
**FREQ:** frequency  
**Gy:** gray  
**ID:** identification  
**MP:** monitoring post  
**SAS:** Statistical Analysis System  
**SE:** standard error  
**Sv:** Sievert

*Edited by G Eysenbach; submitted 27.11.15; peer-reviewed by K Zhao, T Vaughan, S Okun; comments to author 24.01.16; revised version received 26.09.16; accepted 23.10.16; published 25.11.16.*

*Please cite as:*

*Nishimoto N, Ota M, Yagahara A, Ogasawara K*

*Estimating the Duration of Public Concern After the Fukushima Dai-ichi Nuclear Power Station Accident From the Occurrence of Radiation Exposure-Related Terms on Twitter: A Retrospective Data Analysis*

*JMIR Public Health Surveill 2016;2(2):e168*

*URL: <http://publichealth.jmir.org/2016/2/e168/>*

*doi: [10.2196/publichealth.5384](https://doi.org/10.2196/publichealth.5384)*

*PMID: [27888168](https://pubmed.ncbi.nlm.nih.gov/27888168/)*

©Naoki Nishimoto, Mizuki Ota, Ayako Yagahara, Katsuhiko Ogasawara. Originally published in JMIR Public Health and Surveillance (<http://publichealth.jmir.org>), 25.11.2016. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Public Health and Surveillance, is properly cited. The complete bibliographic information, a link to the original publication on <http://publichealth.jmir.org>, as well as this copyright and license information must be included.

Original Paper

# E-Cigarette Social Media Messages: A Text Mining Analysis of Marketing and Consumer Conversations on Twitter

Allison J Lazard<sup>1,2\*</sup>, PhD; Adam J Saffer<sup>1\*</sup>, PhD; Gary B Wilcox<sup>3\*</sup>, PhD; Arnold DongWoo Chung<sup>3\*</sup>, MS; Michael S Mackert<sup>3\*</sup>, PhD; Jay M Bernhardt<sup>2\*</sup>, MPH, PhD

<sup>1</sup>School of Media and Journalism, University of North Carolina at Chapel Hill, Chapel Hill, NC, United States

<sup>2</sup>Center for Health Communication, Moody College of Communication, The University of Texas at Austin, Austin, TX, United States

<sup>3</sup>Center for Health Communication, Stan Richards School of Advertising and Public Relations, The University of Texas at Austin, Austin, TX, United States

\* all authors contributed equally

**Corresponding Author:**

Allison J Lazard, PhD

Center for Health Communication

Moody College of Communication

The University of Texas at Austin

CMB 3.120

Austin, TX

United States

Phone: 1 317 331 7807

Fax: 1 919 962 0620

Email: [lazard@unc.edu](mailto:lazard@unc.edu)

## Abstract

**Background:** As the use of electronic cigarettes (e-cigarettes) rises, social media likely influences public awareness and perception of this emerging tobacco product.

**Objective:** This study examined the public conversation on Twitter to determine overarching themes and insights for trending topics from commercial and consumer users.

**Methods:** Text mining uncovered key patterns and important topics for e-cigarettes on Twitter. SAS Text Miner 12.1 software (SAS Institute Inc) was used for descriptive text mining to reveal the primary topics from tweets collected from March 24, 2015, to July 3, 2015, using a Python script in conjunction with Twitter's streaming application programming interface. A total of 18 keywords related to e-cigarettes were used and resulted in a total of 872,544 tweets that were sorted into overarching themes through a text topic node for tweets (126,127) and retweets (114,451) that represented more than 1% of the conversation.

**Results:** While some of the final themes were marketing-focused, many topics represented diverse proponent and user conversations that included discussion of policies, personal experiences, and the differentiation of e-cigarettes from traditional tobacco, often by pointing to the lack of evidence for the harm or risks of e-cigarettes or taking the position that e-cigarettes should be promoted as smoking cessation devices.

**Conclusions:** These findings reveal that unique, large-scale public conversations are occurring on Twitter alongside e-cigarette advertising and promotion. Proponents and users are turning to social media to share knowledge, experience, and questions about e-cigarette use. Future research should focus on these unique conversations to understand how they influence attitudes towards and use of e-cigarettes.

(*JMIR Public Health Surveill* 2016;2(2):e171) doi:[10.2196/publichealth.6551](https://doi.org/10.2196/publichealth.6551)

**KEYWORDS**

e-cigarettes; social media; tweet; Internet



## Introduction

Since the introduction of electronic nicotine delivery systems or electronic cigarettes (e-cigarettes) less than a decade ago, awareness and use of these products has risen dramatically [1-5]. E-cigarettes are battery-powered devices that deliver varying amounts of nicotine, propylene glycol, and flavorings, among other things, through an aerosolized liquid solution (commonly referred to as vaping) [6]. Introduced as a substitution for cigarettes, e-cigarette use has grown rapidly despite little knowledge about the short- and long-term health risks or the broad impact on public health given high levels of dual use of e-cigarettes and conventional cigarettes [6-9]. Concerns for increased e-cigarette use are not limited to dual-users. Lifetime or ever use of e-cigarettes has increased exponentially among adults, and more importantly, those most susceptible to addiction, adolescents and young adults [10-13].

As the prevalence of e-cigarettes increases, researchers must understand how individuals acquire knowledge about these products and consider how the different sources of information might influence decisions of use. Prior research has consistently found that individuals are exposed to, search for, and share information about e-cigarettes on social media platforms like Twitter and Facebook, among others [14-17]. Interestingly, e-cigarette users are “more likely to be exposed to e-cigarette information via TV viewing sites (eg, Hulu), Twitter, Facebook, email, and Internet search engines” than nonusers [15]. Therefore, it is important to understand what information e-cigarette users and potential users are exposed to on social media.

Twitter is particularly unique among social media platforms because users can broadcast their messages in the form of tweets and retweets to large audiences [18-20], often sharing information that may contribute to conceptualizations of new health products and phenomena [17,21,22]. Broadly speaking, there are 3 types of users tweeting about e-cigarettes: individual users representative of the public, e-cigarette proponents, and commercial marketers. Individual users are average people who tweet or retweet messages about e-cigarettes that can directly reach their followers or indirectly reach other users following hashtags or key terms in a tweet. Proponents are a diverse group of representatives of e-cigarette organizations, vaping advocates, and those who identify as e-cigarette users in their online profiles [23]. Notably, proponents have been shown to tweet e-cigarette content 5 times as often as individual users, much of which is positive toward e-cigarette use [23]. Commercial marketers, such as e-cigarette retailers and manufacturers, are those who use Twitter to promote and advertise their products. Individuals can be exposed to information about e-cigarettes from individual, proponent, or commercial sources.

Many studies have found that exposure to e-cigarette messages on Twitter has increased significantly in recent years [24-26]. As of 2016, 1 study found tweets about e-cigarettes have increased 5 times since 2012 [26]. Further investigations into the sources of e-cigarette tweets indicate that most tweets originate from commercial users as marketing or advertising promotional messages. According to Huang and colleagues

[27], 90% of e-cigarette tweets are from commercial users, and Kim et al [24] reported that 93.43% of tweets were for advertising.

However, the data collected for these studies may have had some confounding variables. Huang et al [27] noted that “the time frame during which our data were collected was just prior to the launch of major e-cigarette TV marketing.” Kim et al [24] analyzed data from the same time frame. This may have led to the finding that most e-cigarette tweets come from commercial users. Some evidence is emerging that challenges previously reported prevalence of e-cigarette tweets from commercial users and for advertising. Dai and Hao [28], who analyzed individual users’ organic tweets about e-cigarettes, found that 10.8% supported, 17.7% were against, and 19.4% were neutral toward the use of e-cigarettes [28], indicating there is variance in the public conversation about e-cigarettes on Twitter.

An additional confound is the types of users tweeting about e-cigarettes. To understand individual exposures, it is necessary to reveal the type of accounts generating e-cigarette-related tweets. Researchers must first recognize the prevalence of automated social bot accounts (also referred to as cyborgs) run by third parties to influence and promote e-cigarettes and related products but resembling average users [29,30]. Tweets from bot accounts are reducing the signal-to-noise ratio where individuals’ e-cigarette tweets are far fewer.

For instance, Clark et al [31] looked at a 10% sample of e-cigarette tweets from “Twitter’s garden hose” over a 2-year period and found that 80% of tweets were “automated and promotional.” Thus, it is necessary to increase the signal-to-noise ratio to accurately examine e-cigarette-related Twitter discussions among individuals, and denoising techniques allow researchers to do so [29]. However, Allem and Ferrara [29] cautioned researchers against using “crude” and “blunt” techniques such as removing tweets with links solely relying on community detection or methods solely relying on innocent-by-association paradigms. While denoising techniques are still emerging, this work highlights an important need for awareness of the signal-to-noise issue for surveillance of social media data.

Searching for the signal, themes of e-cigarette tweets have been studied to reveal salient topics and sentiment of publically accessible conversations on Twitter. Using a priori themes, Kavuluru and Sabbir [23] detected largely positive sentiment for themes of flavors, harm reduction, smoke-free aspects, and smoking cessation. Similarly and on a small scale relative to the current data-mining abilities, Myslin and colleagues [32] used a combination of a priori themes and iterative machine learning to identify a preponderance of first- or second-hand experience tweets about tobacco-related themes of hookah, cessation, and pleasure. E-cigarette content was not among the top themes, likely given the newness of these products in 2011-2012; however, a trend of positive sentiment with e-cigarette content was detected (contrary to more negative sentiment with smoking tweets) [32]. Cole-Lewis and colleagues [33,34], also using themes developed from previous research, revealed that advertising/promotion, policy/government, and

health/safety are among the most dominant themes. Much of this content had a positive sentiment skewing favorably toward e-cigarettes [33]. In a purposeful sample specific for smoking cessation, van der Tempel and colleagues [26] found similar dominant themes—marketing, news, personal experiences—from an a priori theme list. Together, the trends in the literature indicate that as e-cigarette content becomes more popular, the majority of this content is positive promotion. However, missing still is large-scale, inductive analysis of the topics and themes of the tweets from all categories of users. Our study presents the topics and themes of tweets about e-cigarettes from individual, proponent, and commercial users.

An accurate understanding of the types of users tweeting, as well as what they are tweeting, about e-cigarettes can provide a better understanding of what individuals are being exposed to on social media. Looking specifically at the nonadvertising tweets in their data, Kim et al [24] found that organic conversations are occurring online among individuals about e-cigarettes. This is particularly important because organic conversations may affect individual exposure. That is, if individuals are exposed to an e-cigarette tweet from another individual they know, they may be more easily persuaded given the relational closeness and potentially stronger levels of source credibility [35,36]. Thus, as the use of e-cigarettes continues to rise, public awareness of these products is likely shaped by the proliferation of messages shared and reshared on social media. To understand this media landscape, this study used a textual analysis method to examine the public conversation on Twitter and determine overarching themes and trending topics from commercial and consumer contributors.

## Methods

### Text Mining and Data Acquisition

This study used a text-mining approach to uncover key patterns and relationships within unstructured data to understand and evaluate information important to the audience. Text mining is the term used to describe either a single process or a collection of processes in which software tools actively engage in the “discovery of new, previously unknown information by automatically extracting information from different written (or text) sources” [37]. Text mining provides an opportunity to uncover key patterns and relationships within both structured and unstructured data and allows researchers to more easily understand and evaluate information important to the audience.

In the area of public health, text mining of social media has been used to detect and track disease outbreaks and estimate the level of public knowledge regarding health issues [38]. Twitter was selected for data collection due to its popularity as a microblog as well as the active nature of its users in sending messages to create conversations regarding the use of new products and related social issues [18,21]. The central challenge of text mining is the analysis of unstructured data in order to extract meaningful associations, trends, and patterns in large amounts of text. The increasing availability and magnitude of unstructured digital data available in social media such as Twitter, Facebook, blogs, and other online environments offers

new opportunities for researchers to investigate social, cultural, and health issues.

The methodology and workflow for this study depended upon a combination of human and technological analysis of the Twitter messages and employed 5 steps. First, the search term and time period were determined. Next, Twitter’s streaming application programming interface (API) was used to acquire the data. Third, researchers cleaned the data and removed duplication and unrecognized characters. Fourth, data were processed using text-mining software and fifth, the findings were interpreted.

### Data Collection

Twitter was selected for data collection due to its popularity as a microblogging service and the active nature of its users in sending messages regarding news and social issues, including health-related issues. Tweets were collected from March 24, 2015, to July 3, 2015, using a Python script in conjunction with Twitter’s streaming API. A total of 18 keywords related to e-cigarettes, vaping, and e-cigarette brands were used (ecigs, ecigarettes, e-cigarettes, electronic cigarettes, vaping, vapestick, ehookah, ejuice, Blu eCigs, E-Swisher, Ezsmoker, Fin, NJOY/NJOY, Smoke Assist, V2 Cigs, MarkTen, Vuse, and Tryst) which resulted in a total of 872,544 tweets and retweets. Tweets and retweets were separated into 2 files for the analysis, a tweet file containing 546,651 entries and a retweet file containing 325,893 entries.

### Data Analysis

Analysis of the textual content of the tweets was conducted using SAS Text Miner version 12.1 (SAS Institute Inc). SAS Text Miner allowed the researchers to parse and extract information from text, filter, and assemble documents into related topics allowing the researchers to discover topics and understand the data. This software was used for descriptive text-mining purposes to uncover the primary topics that were being discussed during the 100 days examined.

Following the collection of the data, the initial step was to extract, clean, and create a dictionary of words from the data using a natural language processor. A node process flow was created in SAS Enterprise Miner Workstation version 12.1 (SAS Institute Inc). It began with a Text Parsing node where each tweet is divided into tokens (terms). Specifically, this includes identifying sentences, determining parts of speech, and stemming words. Words were spell-checked and parsed to identify entities and remove stop words. The identified tokens or terms were listed in a “term by frequency” matrix via a text transformation of the numerical representation of the text using linear algebra-based priority models. To ensure that words that contribute little were not overly emphasized by the algorithm, the following parts of speech were ignored: auxiliary verbs, conjunctions, determiners, interjections, participles, prepositions, and pronouns.

Next in the Text Filter node, terms that appeared in fewer than 10 messages were ignored. The data were filtered using Entropy as the term weight and Log as the frequency weighting. The term filtering alters the term-by-document matrix, which contains the frequency of the occurrence of the term in the

documents as the value of each cell. From this frequency matrix, a weighted term-by-document matrix was generated using software-driven term-weighting techniques. Within the Text Filter, the Filter Viewer was employed to visually inspect the individual terms. Unrecognizable symbols and letter groupings were manually excluded. Next, a check was made of the terms that were ignored to ascertain if any should be included in the analysis. A single author who had knowledge of the subject matter manually excluded irrelevant terms. Finally, the Text Filter node was used to reduce the total number of parsed terms, thereby eliminating extraneous information and retaining the most relevant parts of the text.

The Text Topic node was then employed to combine terms into topic groups. SAS Text Miner uses 2 types of clustering algorithms: expectation maximization (EM) and hierarchical clustering. EM clustering was used because it allows for and automatically selects between 2 versions of the algorithm—1 for small data files (standard) and 1 for larger (scaled) data files. Since there were over 800,000 tweets and retweets, the EM option was preferred. Options were selected within the software to create topic groups to include all topics that contained more

than 1% of the total tweets or retweets. Topics with less than 5467 tweets or 3259 retweets were excluded from the analysis given they represented less than 1% of the data.

Last, the Topic Viewer option in Text Topic was used to further refine and interpret the topic groups. Individual tweets and retweets were reviewed and used to create summaries of each topic group. One author evaluated the results by completing several different iterations of SAS Text Miner, comparing the different results, and selecting what appeared to be the optimum solution after careful inspection of the output. After visual examination of each topic list (9 tweet topics and 14 retweet topics), topics that clearly did not illustrate the main themes were removed to reduce noise. Individual review of the actual topics generated by the software was undertaken to further exclude topics that appeared from automated accounts. This was accomplished by individually reviewing the actual messages from each topic to produce the final grouping of topics for tweets (8 topics) and retweets (5 topics) with the description in [Tables 1 and 2](#). This process aided in noise reduction among the tweets as called for by Allem and Ferrar [29].

**Table 1.** E-cigarette tweets by topic.

SAS Text Miner topic	n	Description of topic	Category
1 +E-cigarette, +smoke, ecigs, +quit, vaping	29,556	Pro-vaping discussion whether e-cigarettes are cessation devices or gateway products for youth to establish nicotine addictions	Proponent and individual user conversation
2 Ejuice, eliquid, vape, vapelife, ecig	24,064	Pro-vaping reactions to e-cigarette policies and questions about e-cigarette health risk claims	Proponent and individual user conversation
3 Vaping, vape, ecigs, vapelife, vapecommunity	21,555	News and updates from the vaping community, primarily through a vaping blog's daily updates	Proponent and individual user conversation
4 Vapeporn, +tree, vape, vaping, reddit	16,414	Vaping advocacy encouraging the uptake of vaping and tips for users	Proponent and individual user conversation
5 Electronic, +cigarette, +electronic cigarette, +employee, +relieve	12,694	Employees may use e-cigarettes at work as relief from smoking bans	Proponent and individual user conversation
6 +Juice, vapejuice, +vaporizer, +vapor, vape	9092	Promotion for e-juice	Marketing/advertising
7 +Well price, +vapour, +price, +good,	6860	Price promotion for e-cigarettes	Marketing/advertising
8 +Win, cigbuyer, vapegiveaway, +enter, +sampler	5892	Promotion for an e-liquid give-away	Marketing/advertising

**Table 2.** E-cigarette retweets by topic.

SAS Text Miner topic	n	Description of topic	Category
1 Vapinxsmoker, vaping, ecigs, +smoke, +e-cigarette	42,430	Discussion of policies banning e-cigarette use, e-cigarettes as (or not as) smoking cessation devices, and differentiation of e-cigarettes from traditional cigarettes	Proponent and individual user conversation
2 Vape, ejuice, ecigs, +win, vaping	33,767	Promotions for an e-juice give-away	Marketing/advertising
3 Amp, +win, +follow, +chance, http	18,305	Promotions for a chance to win an e-cigarette	Marketing/advertising
4 Vape, realdonnadevane, vaping, granny, ur	13,752	E-cigarette blogger suggests ways to "grow your #Vape business"	Marketing/advertising
5 +Starter, hookahcoals, http://t.co/5amkbtyrp3, newyork, eastcoast	6197	Promotions for e-hookah starter kits	Marketing/advertising

## Results

### Overview

Of the 872,544 tweets and retweets captured from March to July 2015, 240,578 were included in the final topic groups that each represented more than 1% of the conversation on Twitter during this time period for tweets and retweets. These were divided among 126,127 tweets sorted into 8 unique topics, shown in Table 1, and 114,451 retweets in 5 topics, shown in Table 2. Each table shows the topics featuring the prevalent keywords as generated by SAS Text Miner, number of tweets per topic, a description of the topic, and whether the topic represents proponent and individual user conversations or marketing/advertising. All topics with diverse tweets not directly linked to marketing were labeled proponent and individual user conversations. Topics dominated by promotions or branded persuasive appeals were categorized as marketing/advertising; often a single, repeated tweet comprised the entire topic for marketing/advertising.

### Tweet Topics

Of the 8 topics generated from the tweets, the top 5 topics, determined by number of tweets contained in the topic, included diverse proponent and individual user conversations. The most popular topic in the analysis represented a diverse public conversation that covered whether e-cigarettes are cessation devices or gateway products to get youth addicted to nicotine. This topic was dominated by pro-e-cigarette content and included comments that e-cigarettes may help people quit smoking, that the rise in e-cigarette use among adolescents may be deterring them from traditional tobacco experimentation, how e-cigarette bans may have unintended consequences, and claims and questions about whether the science showing the risk of e-cigarettes is flawed. Anti-e-cigarette tweets within this topic were limited but did include warnings for adolescent use.

The second most popular topic was a proponent and individual user conversation in reaction to e-cigarette bans and proposed taxes along with further questioning of whether there is evidence to support health risk claims about e-cigarette use. This topic represents a conversation around efforts to differentiate e-cigarettes from traditional tobacco products by pointing to the lack of evidence for the harm or risks of e-cigarettes along with the position that e-cigarettes should be promoted as smoking cessation devices, aside from the youth discussion above, by generally taking the stance that the use of e-cigarettes can save lives. Additionally, concerns expressed that products would be too expensive or unavailable because of regulation were also common.

The third topic from the tweets contained e-cigarette news and updates from proponents in the vaping community, although this community was dominated by one pro-e-cigarette news outlet. Many of these tweets were generated through the “share” option from the daily vaping news website and covered a range of topics that included coverage of policies (eg, bans, taxes), promotions from the organization, product reviews, tips for social and culture practices of vaping, and diverse articles that highlight e-cigarettes as cessation devices, along with the uncertainty of risk with e-cigarette use (eg, “Is a daily dose of

nicotine as benign as coffee?”). This topic also included notices of new products available for sale from a variety of distributors.

The fourth topic consisted of vaping advocacy comments from a variety of proponent and individual user angles. Tweets included tips for e-cigarette users, discussions of flavors, encouragement for expanding the social practices of vaping (via new people and new places), and information about the use of specific devices as well as using devices for marijuana consumption. The fifth proponent and individual user topic consisted of a discussion about how employees may use e-cigarettes as a way to find relief from smoking bans at workplaces, as well as providing information about e-cigarettes and vaping.

Topics 6 through 8, the least populated topics, contained marketing promotions from 3 unique vendors. Each of these topics contained a single repeated tweet or tweets with only slight variations. Topic 6 contained just over 9000 original tweets that promoted a single distributor’s e-juice in a variety of flavors, such as kettle corn, grape, vanilla, and menthol. Topic 7 was entirely the repetition of one price promotion tweet from a different distributor, merely mentioning they had the best prices. Topic 8 was similarly the repeat of a single tweet; the tweet was a give-away promotion for an e-liquid sampler from a third distributor. The sampler promotion contained four 30 mL bottles in flavors *gravel pit*, *lime cola*, *strawberry blonde*, and *trail mix*.

### Retweet Topics

The 5 retweet topics consisted of 1 proponent and individual user conversation topic and 4 marketing/advertising topics. The most popular topic—the proponent and individual user conversation—contained references to policy bans for using e-cigarettes in public places and raising age restrictions for the legal purchase of tobacco products; comments about the likelihood that e-cigarette use does (or does not) lead to smoking cessation, with a heavy emphasis on how switching to e-cigarettes may not help users quit; and arguments for the differentiation between vaping and smoking, often framed as a pro-vaping argument focused on the reduced risk of e-cigarette use compared to smoking cigarettes.

The remaining topics, which make up over 60% of the retweeted content, were comprised exclusively of 4 unique messages retweeted over 70,000 times. Most (3 of the 4) retweeted messages were promotions to win a free bottle of e-juice, a variable wattage mod style e-cigarette, or an e-hookah starter kit. The remaining marketing/advertising topic of retweeted messages was a promotion for a single pro-e-cigarette blogger as an endorser and consultant for vape businesses.

## Discussion

### Principal Findings

As the use of e-cigarettes continues to rise, public awareness and perception of these products are likely shaped by the proliferation of messages shared and reshared on social media [26]. This study examined the public conversation on Twitter to determine overarching themes and trending topics. Topics found in this study included whether e-cigarettes are cessation



devices or gateway products for tobacco addiction, how e-cigarettes differ from traditional cigarettes, reactions to e-cigarette policies and health risk claims, news and updates from vaping communities, use of e-cigarettes where smoking bans exist, and a variety of marketing product promotions and giveaways. While past research has found a preponderance of marketing and advertising dominating the content on Twitter [24,27], this study revealed that proponents and individual users are participating in public conversations about e-cigarettes at a much larger scale than previously suggested. However, the diversity in content did not reveal a conversation with diverse perspectives. Individual user and proponent tweets intertwined with marketing messages, which still have a strong presence, to present a rather unbalanced, likely proponent-driven and perpetuated conversation about e-cigarettes use, norms, and policy.

In contrast to earlier studies, this study is the first to indicate that public conversations, from a mix of individual users and proponents, are now dominating the trending topics on Twitter for e-cigarettes, even with the inclusion of commercial activity [26]. These topics, with over 800,000 tweets generated in a 100-day window, provide rich insights into salient issues for Twitter users, especially for those who support the use of e-cigarettes. Advocates for e-cigarettes have taken to Twitter to share their thoughts and opinions, contributing to an unbalanced, likely proponent-driven conversation about e-cigarettes that heavily favors pro-e-cigarette arguments. Notably, although the most popular topics were labeled proponent and individual users conversations, these were not free of industry influence via a mixture of marketing tweets within as well as messages clearly influenced by marketing strategies from users and industry members alike. Content from e-cigarette proponents reflects the indirect marketing influence in these conversations.

It is perhaps not surprising that those passionate enough to tweet are talking about the benefits of e-cigarette use. Similar trends of Twitter conversations dominated with pro-e-cigarette content have been detected in response to e-cigarette educational campaigns and announcements of e-cigarette regulations [39,40]. Our findings, which mirror these trends of countercampaigns and antipolicy Twitter bombing to flood conversations with one perspective [39,40], highlight the pressing need for public health professionals to engage the public on social media. This finding reveals perhaps an even larger concern for all public health professionals: each person who goes online to do a little research about e-cigarettes is going to encounter a tilted conversation encouraging e-cigarette use, promoting vaping as a socially acceptable practice for all ages, discrediting scientific evidence for health risks, and rallying around the idea that e-cigarettes should largely be outside the bounds of policy. Thinking about how public health advocates can either more actively engage in this conversation or encourage a broader range of the public, inclusive of those with neutral or anti-e-cigarette positions, to post is necessary to create a more balanced conversation.

Marketing and advertising still have a strong presence on Twitter; however, our results only partially support what others have found. Previous studies have suggested that marketing

content saturates over 90% of the information about e-cigarettes on Twitter [24,27]. Although still present in this sample, this analysis revealed a much smaller proportion of marketing/advertising content as the most popular topics, especially for tweets. The fact that the explicitly marketing messages fall behind the individual user tweets indicates that they do not spread or influence the conversation as much as has been shown in previous years.

Marketing messages do spread when specific promotions are retweeted verbatim. A limited number of 140-character-or-less messages that focused on promotions and giveaways proliferated rapidly through retweets during the 100 days in this study. These retweets can be interpreted as public reactions [18], demonstrating that some attention toward promotion of regulations for e-cigarette prices and give-away promotions online is warranted. Taken together, however, explicit marketing content, although retweeted more, was a relatively small piece of the online conversation. While attention to policy and regulations for marketing content might be worthwhile, the more important public health effort is likely to focus on engaging in the conversation to create a more balanced perspective available to Twitter users.

### Limitations

As with all social media research and analyses, there are several limitations to this study. While an analysis that captured a wide breadth of tweets and retweets allowed for insights about large-scale theme and topics, this does not represent the exposure for all Twitter users. Individuals customize their Twitter experiences by following accounts, thus not all users would experience the content shared on social media in the same way. Additionally, although insights for individual user and proponent opinions can be made from content shared in the topics, this analysis does not reveal the impact of the tweets or retweets on perception and attitudes toward e-cigarettes from users who see this content. Lastly, no automated denoising technique was applied prior to text mining the tweets captured in this study; we took steps to reduce noise manually. As emerging techniques become more reliable [29], future studies should consider applying automated denoising techniques before analysis.

### Conclusions

As the prevalence of e-cigarettes continues to rise, it is important to know what messages about these products are potentially influencing consumer attitudes and use. This study is the first to uncover trending themes and topics from large-scale public conversations on social media.

While e-cigarette brands and distributors continue to use social media for e-cigarette marketing and promotion, these findings reveal that unique, large-scale consumer conversations are taking place on Twitter. Individuals are turning to social media to participate in discussions about policies, personal experiences, and the differentiation of e-cigarettes from traditional tobacco. Public health advocates should actively participate on social media to balance the conversation, and future research should investigate how these unique conversations influence attitudes toward and use of e-cigarettes.



## Conflicts of Interest

None declared.

## References

1. 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](https://doi.org/10.2105/AJPH.2011.300526)] [Medline: [22813087](https://pubmed.ncbi.nlm.nih.gov/22813087/)]
2. Loukas A, Batanova M, Fernandez A, Agarwal D. Changes in use of cigarettes and non-cigarette alternative products among college students. *Addict Behav* 2015 Oct;49:46-51. [doi: [10.1016/j.addbeh.2015.05.005](https://doi.org/10.1016/j.addbeh.2015.05.005)] [Medline: [26046401](https://pubmed.ncbi.nlm.nih.gov/26046401/)]
3. King BA, Patel R, Nguyen K, Dube SR. Trends in awareness and use of electronic cigarettes among US adults, 2010-2013. *Nicotine Tob Res* 2015;17(2):219-227. [doi: [10.1093/ntr/ntu191](https://doi.org/10.1093/ntr/ntu191)]
4. Ramo DE, Young-Wolff KC, Prochaska JJ. Prevalence and correlates of electronic-cigarette use in young adults: findings from three studies over five years. *Addict Behav* 2015 Feb;41:142-147 [FREE Full text] [doi: [10.1016/j.addbeh.2014.10.019](https://doi.org/10.1016/j.addbeh.2014.10.019)] [Medline: [25452058](https://pubmed.ncbi.nlm.nih.gov/25452058/)]
5. Adkison SE, O'Connor RJ, Bansal-Travers M, Hyland A, Borland R, Yong H, et al. Electronic nicotine delivery systems: international tobacco control four-country survey. *Am J Prev Med* 2013 Mar;44(3):207-215 [FREE Full text] [doi: [10.1016/j.amepre.2012.10.018](https://doi.org/10.1016/j.amepre.2012.10.018)] [Medline: [23415116](https://pubmed.ncbi.nlm.nih.gov/23415116/)]
6. Grana R, Benowitz N, Glantz SA. E-cigarettes: a scientific review. *Circulation* 2014 May 13;129(19):1972-1986 [FREE Full text] [doi: [10.1161/CIRCULATIONAHA.114.007667](https://doi.org/10.1161/CIRCULATIONAHA.114.007667)] [Medline: [24821826](https://pubmed.ncbi.nlm.nih.gov/24821826/)]
7. Cobb NK, Byron MJ, Abrams DB, Shields PG. Novel nicotine delivery systems and public health: the rise of the “e-cigarette”. *Am J Public Health* 2010 Dec;100(12):2340-2342. [doi: [10.2105/AJPH.2010.199281](https://doi.org/10.2105/AJPH.2010.199281)] [Medline: [21068414](https://pubmed.ncbi.nlm.nih.gov/21068414/)]
8. Williams M, Villarreal A, Bozhilov K, Lin S, Talbot P. Metal and silicate particles including nanoparticles are present in electronic cigarette cartomizer fluid and aerosol. *PLoS One* 2013;8(3) [FREE Full text] [doi: [10.1371/journal.pone.0057987](https://doi.org/10.1371/journal.pone.0057987)] [Medline: [23526962](https://pubmed.ncbi.nlm.nih.gov/23526962/)]
9. Kosmider L, Sobczak A, Fik M, Knysak J, Zaciera M, Kurek J, et al. Carbonyl compounds in electronic cigarette vapors: effects of nicotine solvent and battery output voltage. *Nicotine Tob Res* 2014 Oct;16(10):1319-1326 [FREE Full text] [doi: [10.1093/ntr/ntu078](https://doi.org/10.1093/ntr/ntu078)] [Medline: [24832759](https://pubmed.ncbi.nlm.nih.gov/24832759/)]
10. King BA, Patel R, Nguyen KH, Dube SR. Trends in awareness and use of electronic cigarettes among US adults, 2010-2013. *Nicotine Tob Res* 2015 Feb;17(2):219-227 [FREE Full text] [doi: [10.1093/ntr/ntu191](https://doi.org/10.1093/ntr/ntu191)] [Medline: [25239961](https://pubmed.ncbi.nlm.nih.gov/25239961/)]
11. Centers for Disease Control and Prevention. Tobacco Use Among Middle and High School Students—United States, 2011-2015. *MMWR* 2016 Apr 16;65(14):361-367 [FREE Full text]
12. Regan A, Promoff G, Dube S, Arrazola R. Electronic nicotine delivery systems: adult use/awareness of the “e-cigarette” in the USA. *Tob Control* 2013;22(1):19-23. [doi: [10.1136/tobaccocontrol-2011-050044](https://doi.org/10.1136/tobaccocontrol-2011-050044)]
13. Sutfin EL, McCoy TP, Morrell HE, Hoepfner BB, Wolfson M. Electronic cigarette use by college students. *Drug Alcohol Depend* 2013 Aug 1;131(3):214-221 [FREE Full text] [doi: [10.1016/j.drugalcdep.2013.05.001](https://doi.org/10.1016/j.drugalcdep.2013.05.001)] [Medline: [23746429](https://pubmed.ncbi.nlm.nih.gov/23746429/)]
14. Dawkins L, Turner J, Roberts A, Soar K. “Vaping” profiles and preferences: an online survey of electronic cigarette users. *Addict* 2013;108(6):1115-1125. [doi: [10.1111/add.12150](https://doi.org/10.1111/add.12150)]
15. Emery S, Vera L, Huang J, Szczyepka G. Wanna know about vaping? Patterns of message exposure, seeking and sharing information about e-cigarettes across media platforms. *Tob Control* 2014;23(suppl 3):iii17-iii25. [doi: [10.1136/tobaccocontrol-2014-051648](https://doi.org/10.1136/tobaccocontrol-2014-051648)]
16. Pepper J, Brewer N. Electronic nicotine delivery system (electronic cigarette) awareness, use, reactions and beliefs: a systematic review. *Tob Control* 2014 Sep;23(5):375-384 [FREE Full text] [doi: [10.1136/tobaccocontrol-2013-051122](https://doi.org/10.1136/tobaccocontrol-2013-051122)] [Medline: [24259045](https://pubmed.ncbi.nlm.nih.gov/24259045/)]
17. Chu K, Sidhu A, Valente T. Electronic cigarette marketing online: a multi-site, multi-product comparison. *JMIR Public Health Surveill* 2015;1(2):e11 [FREE Full text] [doi: [10.2196/publichealth.4777](https://doi.org/10.2196/publichealth.4777)] [Medline: [27227129](https://pubmed.ncbi.nlm.nih.gov/27227129/)]
18. Boyd D, Golder S, Lotan G. Tweet, tweet, retweet: Conversational aspects of retweeting on twitter. 2010 Presented at: System Sciences (HICSS) 43rd Hawaii International Conference; 2010; Hawaii.
19. Chen G. Tweet this: A uses and gratifications perspective on how active Twitter use gratifies a need to connect with others. *Computers in Human Behavior* 2011 Mar;27(2):755-762. [doi: [10.1016/j.chb.2010.10.023](https://doi.org/10.1016/j.chb.2010.10.023)]
20. Wu S, Hofman J, Mason W, Watts D. Who says what to whom on twitter. 2011 Presented at: Proceedings of the 20th International Conference on World Wide Web; 2011; Hyderabad, India.
21. Scanfeld D, Scanfeld V, Larson E. Dissemination of health information through social networks: twitter and antibiotics. *Am J Infect Control* 2010 Apr;38(3):182-188 [FREE Full text] [doi: [10.1016/j.ajic.2009.11.004](https://doi.org/10.1016/j.ajic.2009.11.004)] [Medline: [20347636](https://pubmed.ncbi.nlm.nih.gov/20347636/)]
22. Radzikowski J, Stefanidis A, Jacobsen K, Croitoru A, Crooks A, Delamater P. The measles vaccination narrative in Twitter: a quantitative analysis. *JMIR Public Health Surveill* 2016;2(1):e1 [FREE Full text] [doi: [10.2196/publichealth.5059](https://doi.org/10.2196/publichealth.5059)] [Medline: [27227144](https://pubmed.ncbi.nlm.nih.gov/27227144/)]

23. Kavuluru R, Sabbir AK. Toward automated e-cigarette surveillance: Spotting e-cigarette proponents on Twitter. *J Biomed Inform* 2016 Jun;61:19-26. [doi: [10.1016/j.jbi.2016.03.006](https://doi.org/10.1016/j.jbi.2016.03.006)] [Medline: [26975599](https://pubmed.ncbi.nlm.nih.gov/26975599/)]
24. Kim A, Hopper T, Simpson S, Nonnemaker J, Lieberman AJ, Hansen H, et al. Using Twitter data to gain insights into e-cigarette marketing and locations of use: an infoveillance study. *J Med Internet Res* 2015 Nov 06;17(11):e251 [FREE Full text] [doi: [10.2196/jmir.4466](https://doi.org/10.2196/jmir.4466)] [Medline: [26545927](https://pubmed.ncbi.nlm.nih.gov/26545927/)]
25. Chu K, Unger J, Allem JP, Pattaroyo M, Soto D, Cruz TB, et al. Diffusion of messages from an electronic cigarette brand to potential users through Twitter. *PLoS One* 2015;10(12) [FREE Full text] [doi: [10.1371/journal.pone.0145387](https://doi.org/10.1371/journal.pone.0145387)] [Medline: [26684746](https://pubmed.ncbi.nlm.nih.gov/26684746/)]
26. van der Tempel J, Noormohamed A, Schwartz R, Norman C, Malas M, Zawertailo L. Vape, quit, tweet? Electronic cigarettes and smoking cessation on Twitter. *Int J Public Health* 2016 Mar;61(2):249-256. [doi: [10.1007/s00038-016-0791-2](https://doi.org/10.1007/s00038-016-0791-2)] [Medline: [26841895](https://pubmed.ncbi.nlm.nih.gov/26841895/)]
27. 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](https://doi.org/10.1136/tobaccocontrol-2014-051551)] [Medline: [24935894](https://pubmed.ncbi.nlm.nih.gov/24935894/)]
28. Dai H, Hao J. Mining social media data for opinion polarities about electronic cigarettes. *Tob Control* 2016. [doi: [10.1136/tobaccocontrol-2015-052818](https://doi.org/10.1136/tobaccocontrol-2015-052818)]
29. Allem JP, Ferrara E. The importance of debiasing social media data to better understand e-cigarette-related attitudes and behaviors. *J Med Internet Res* 2016 Aug 09;18(8):e219 [FREE Full text] [doi: [10.2196/jmir.6185](https://doi.org/10.2196/jmir.6185)] [Medline: [27507563](https://pubmed.ncbi.nlm.nih.gov/27507563/)]
30. Ferrara E, Varol O, Davis C, Menczer F, Flammini A. The rise of social bots. *Commun ACM* 2016 Jun 24;59(7):96-104. [doi: [10.1145/2818717](https://doi.org/10.1145/2818717)]
31. Clark EM, Jones CA, Williams JR, Kurti AN, Norotsky MC, Danforth CM, et al. Vaporous marketing: uncovering pervasive electronic cigarette advertisements on Twitter. *PLoS One* 2016;11(7):e0157304 [FREE Full text] [doi: [10.1371/journal.pone.0157304](https://doi.org/10.1371/journal.pone.0157304)] [Medline: [27410031](https://pubmed.ncbi.nlm.nih.gov/27410031/)]
32. 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 Aug 29;15(8):e174 [FREE Full text] [doi: [10.2196/jmir.2534](https://doi.org/10.2196/jmir.2534)] [Medline: [23989137](https://pubmed.ncbi.nlm.nih.gov/23989137/)]
33. Cole-Lewis H, Pugatch J, Sanders A, Varghese A, Posada S, Yun C, et al. Social listening: a content analysis of e-cigarette discussions on Twitter. *J Med Internet Res* 2015 Oct 27;17(10):e243 [FREE Full text] [doi: [10.2196/jmir.4969](https://doi.org/10.2196/jmir.4969)] [Medline: [26508089](https://pubmed.ncbi.nlm.nih.gov/26508089/)]
34. Cole-Lewis H, Varghese A, Sanders A, Schwarz M, Pugatch J, Augustson E. Assessing electronic cigarette-related Tweets for sentiment and content using supervised machine learning. *J Med Internet Res* 2015 Aug 25;17(8):e208 [FREE Full text] [doi: [10.2196/jmir.4392](https://doi.org/10.2196/jmir.4392)] [Medline: [26307512](https://pubmed.ncbi.nlm.nih.gov/26307512/)]
35. Westerman D, Spence P, Van Der Heide B. Social media as information source: recency of updates and credibility of information. *J Comput-Mediat Comm* 2013 Nov 08;19(2):171-183. [doi: [10.1111/jcc4.12041](https://doi.org/10.1111/jcc4.12041)]
36. Dubois E, Gaffney D. The multiple facets of influence: identifying political influentials and opinion leaders on Twitter. *American Behavioral Scientist* 2014 Apr 08;58(10):1260-1277. [doi: [10.1177/0002764214527088](https://doi.org/10.1177/0002764214527088)]
37. Ampofo L, Collister S, O'Loughlin B, Chadwick A. Text mining and social media: When quantitative meets qualitative and software meets people. In: Halfpenny P, Proctor R, editors. *Innovations in Digital Research Methods*. Thousand Oaks: Sage Publications Inc; 2015:161-192.
38. Broniatowski D, Paul M, Dredze M. National and local influenza surveillance through Twitter: an analysis of the 2012-2013 influenza epidemic. *PLoS One* 2013;8(12):e83672 [FREE Full text] [doi: [10.1371/journal.pone.0083672](https://doi.org/10.1371/journal.pone.0083672)] [Medline: [24349542](https://pubmed.ncbi.nlm.nih.gov/24349542/)]
39. Harris JK, 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 Oct 16;16(10):e238 [FREE Full text] [doi: [10.2196/jmir.3622](https://doi.org/10.2196/jmir.3622)] [Medline: [25320863](https://pubmed.ncbi.nlm.nih.gov/25320863/)]
40. Allem J, Escobedo P, Chu K, Soto D, Cruz T, Unger J. Campaigns and counter campaigns: reactions on Twitter to e-cigarette education. *Tob Control* 2016. [doi: [10.1136/tobaccocontrol-2015-052757](https://doi.org/10.1136/tobaccocontrol-2015-052757)]

---

## Abbreviations

- API:** application programming interface  
**e-cigarette:** electronic cigarette  
**EM:** expectation maximization
-

*Edited by T Sanchez; submitted 26.08.16; peer-reviewed by JP Allem, H Yang; comments to author 19.09.16; revised version received 31.10.16; accepted 16.11.16; published 12.12.16.*

*Please cite as:*

*Lazard AJ, Saffer AJ, Wilcox GB, Chung AD, Mackert MS, Bernhardt JM*

*E-Cigarette Social Media Messages: A Text Mining Analysis of Marketing and Consumer Conversations on Twitter*

*JMIR Public Health Surveill 2016;2(2):e171*

*URL: <http://publichealth.jmir.org/2016/2/e171/>*

*doi: [10.2196/publichealth.6551](https://doi.org/10.2196/publichealth.6551)*

*PMID: [27956376](https://pubmed.ncbi.nlm.nih.gov/27956376/)*

©Allison J Lazard, Adam J Saffer, Gary B Wilcox, Arnold DongWoo Chung, Michael S Mackert, Jay M Bernhardt. Originally published in JMIR Public Health and Surveillance (<http://publichealth.jmir.org>), 12.12.2016. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Public Health and Surveillance, is properly cited. The complete bibliographic information, a link to the original publication on <http://publichealth.jmir.org>, as well as this copyright and license information must be included.

Original Paper

# Disease Monitoring and Health Campaign Evaluation Using Google Search Activities for HIV and AIDS, Stroke, Colorectal Cancer, and Marijuana Use in Canada: A Retrospective Observational Study

---

Rebecca Ling<sup>1</sup>, BSc Hons; Joon Lee<sup>1</sup>, PhD

School of Public Health and Health Systems, Faculty of Applied Health Sciences, University of Waterloo, Waterloo, ON, Canada

**Corresponding Author:**

Joon Lee, PhD

School of Public Health and Health Systems

Faculty of Applied Health Sciences

University of Waterloo

200 University Avenue West

Waterloo, ON, N2L 3G1

Canada

Phone: 1 519 888 4567 ext 31567

Fax: 1 519 746 6776

Email: [joon.lee@uwaterloo.ca](mailto:joon.lee@uwaterloo.ca)

---

## Abstract

**Background:** Infodemiology can offer practical and feasible health research applications through the practice of studying information available on the Web. Google Trends provides publicly accessible information regarding search behaviors in a population, which may be studied and used for health campaign evaluation and disease monitoring. Additional studies examining the use and effectiveness of Google Trends for these purposes remain warranted.

**Objective:** The objective of our study was to explore the use of infodemiology in the context of health campaign evaluation and chronic disease monitoring. It was hypothesized that following a launch of a campaign, there would be an increase in information seeking behavior on the Web. Second, increasing and decreasing disease patterns in a population would be associated with search activity patterns. This study examined 4 different diseases: human immunodeficiency virus (HIV) infection, stroke, colorectal cancer, and marijuana use.

**Methods:** Using Google Trends, relative search volume data were collected throughout the period of February 2004 to January 2015. Campaign information and disease statistics were obtained from governmental publications. Search activity trends were graphed and assessed with disease trends and the campaign interval. Pearson product correlation statistics and joinpoint methodology analyses were used to determine significance.

**Results:** Disease patterns and online activity across all 4 diseases were significantly correlated: HIV infection ( $r=.36$ ,  $P<.001$ ), stroke ( $r=.40$ ,  $P<.001$ ), colorectal cancer ( $r=-.41$ ,  $P<.001$ ), and substance use ( $r=.64$ ,  $P<.001$ ). Visual inspection and the joinpoint analysis showed significant correlations for the campaigns on colorectal cancer and marijuana use in stimulating search activity. No significant correlations were observed for the campaigns on stroke and HIV regarding search activity.

**Conclusions:** The use of infoveillance shows promise as an alternative and inexpensive solution to disease surveillance and health campaign evaluation. Further research is needed to understand Google Trends as a valid and reliable tool for health research.

(*JMIR Public Health Surveill* 2016;2(2):e156) doi:[10.2196/publichealth.6504](https://doi.org/10.2196/publichealth.6504)

---

## KEYWORDS

public health informatics; Internet; information seeking behavior

## Introduction

With an increasing number of people using the World Wide Web, their activities generate “big data” and provide meaningful research in infodemiology, which is the study of patterns and determinants of information on the Web or in a population with the purpose to inform public health and public policy [1]. Infodemiology is of two types, supply based and demand based, which are contrasted by their focus on the provision or uptake of information, respectively. Supply-based infodemiology examines the amount and quality of information available on the Web, such as the number of websites, media reports, and blogs related to vaccinations as well as the level of quality (eg, positive or negative portrayals of vaccines) [2]. On the contrary, demand-based infodemiology examines the patterns of people’s search behaviors for seeking information on the Web, such as the increase in search activities in light of related news reports [2]. Furthermore, a specific branch of demand-based infodemiology known as infoveillance is the observation of information seeking behaviors on the Web. Notably, the use of Web-based data heavily relies on crowdsourcing from the public and their engagement online. Similarly, other applications such as digital disease detection and surveillance work on a similar basis to effectively utilize the continuous generation of Web-based data from this growing digital age [3,4]. However, in this study, the applications of infoveillance in health research are specifically evaluated.

Infoveillance has proven to be successful in predicting infectious disease outbreaks, spawning the development of Google Flu Trends [2,3]. Google Flu Trends leverages the principle that changes in information and communication patterns portray early “symptoms” of a disease emerging in a population [1]. There is strong evidence regarding the predictive utility of search activity and infectious disease outbreak from digital epidemiologic studies. Milinovich et al [4] identified 17 infectious diseases that were positively correlated with search activity trends on Google, with a majority being vaccine-preventable, vector-borne, and sexually transmitted diseases. Most recently, the Colombian Zika virus outbreak showed positive association with search activities made on Google [5]. Current research in infoveillance has primarily focused on infectious diseases, but less is known about its utility pertaining to chronic diseases [6]. In a systematic review looking at the applications of Google Trends in health research, among the 70 included studies, most studies focused on general population surveillance (33%), followed by infectious diseases (27%), mental health and addictions (24%), and noncommunicable diseases (16%) [6].

With the growing digital era, many people now turn to the Web to learn about disease symptoms, diagnosis, and treatments, such as various cancers [7]. One study surveyed more than 12,000 individuals from 12 countries and reported that more than 45% of respondents sought health information on the Web to self-diagnose their condition [8]. Furthermore, these Web-based information seeking behaviors may be more common among diseases that are highly stigmatized in a society, owing to the perception of the disease, to deter from seeking professional help [9]. Thus, such Web-based search activities

can provide valuable insights about the prevalence and distribution of chronic diseases. Disease search trends have been studied for breast cancer [10,11], lung cancer [12], status epilepticus [7], transient ischemic attack [13,14], and kidney stones [15]. Regarding stigmatized diseases, research is available on dementia [16], mental health [9], and suicide deaths [17,18]. Overall, these studies reported moderate to strong degree of associations, suggesting a potential value in applying infoveillance to chronic diseases. Foroughi et al [19] found that cancer-related terms were associated with Web search activity for the burden of cancers, such as cancer incidence and mortality, in Australia, New Zealand, the United Kingdom, Canada, and the United States. Also, an American study showed associations between suicide rates and Google search trends relating to “how to suicide,” “commit suicide,” and “suicide prevention” among all 50 states in 2009 [18]. By applying key search terms associated with the disease outcome, search activity trends may be used to portray further insights about the population health and behaviors. However, in comparison with the numerous studies done on infectious diseases, further research is needed to validate the use of infoveillance for chronic diseases.

Second, infoveillance has also been used to monitor and track the success of marketing campaigns measured by the generation of interest and activity of a population observed on the Internet [2]. Traditional epidemiologic tools, such as nationwide surveys, cohort studies, or registries, are often labor and resource intensive. Because of these factors, it could be very difficult and impractical to observe population trends happening before, during, and after the campaign period. Thus, infoveillance offers an advantage and practicality over traditional methods in campaign evaluation. Currently, few studies have investigated public health campaigns and their effects on Web searches. The most commonly studied campaign is the Breast Cancer Awareness Month campaign established in the United States. These studies reported that the Breast Cancer Awareness Month campaign was effective in triggering search activity [10,11,20]. Swenson and Lindblom [21] analyzed Google Trends for “breast cancer,” “colon cancer,” and “cervical cancer” and reported that cancer awareness month campaigns, specifically breast cancer and colon cancer, increased public awareness as shown by increased Web search activities. In the Republic of Ireland, Davis et al [22] found a direct association between media campaigns for erectile dysfunction and related search activity patterns. Despite limited research on the influence of health campaigns on search activity, other studies showed that media events including news reports and announcements regarding public figures also prompted health seeking behaviors [22-24]. These findings suggest potential value for assessing search activity trends; however, further research is needed in this area to evaluate the effectiveness of infoveillance for monitoring health campaigns.

Thus, the aim of this observational study was to explore the applications of infoveillance in information seeking behavior for human immunodeficiency virus (HIV) and acquired immunodeficiency syndrome (AIDS), stroke, colorectal cancer, and marijuana use. Using Google Trends, these search activities were assessed against census data and public health campaigns



to examine their relationships. To the best of our knowledge, these diseases have not been studied, particularly in the Canadian context. We first hypothesized that search trends are associated with the disease patterns of a population. Second, we hypothesized that the launch of the health campaign stimulates Google search activity, shown on Google Trends. The outcomes of this research provide new insights for public health professionals and contribute to further understanding of infodemiology in health research.

## Methods

### Study Design

In this retrospective study, search activity on Google was examined for colorectal cancer, HIV and AIDS, stroke, and marijuana use from 2004 to 2015 in Canada. Google Trends provided relative search volumes (RSVs) for particular search queries. First, to study the effectiveness of infoveillance in chronic disease monitoring, the search activities were compared against disease prevalence on an annual basis between 2004 and 2015. Second, to investigate the application of infoveillance in health campaign evaluation, the levels of search activity before, during, and after the campaign were analyzed.

### Selection of Health Campaigns

Health campaigns were sought through peer-reviewed sources and gray literature. The search was narrowed to identify campaigns in Canada and campaigns implemented after 2004 because Google Trends provides its data only after this time point. Health campaigns through any medium, such as televised advertisements, program delivery, or pamphlet distribution, were considered in the assessment. However, the minimum data elements required for this study were the campaign duration, frequency, and location. Health campaigns that met these criteria were then screened based on their disease focus. Preference was given to diseases that had not yet been reviewed in current literature, as well as chronic diseases in order to examine the infodemiology applications for chronic disease monitoring. Thus, health campaigns were chosen based on their transparency and availability of public information, which subsequently dictated the 4 diseases studied in this paper.

As a result, this includes the “ColonCancerCheck” campaign led by the Government of Ontario, “End HIV Stigma” campaign led by the Positive Living Society of British Columbia (formerly

known as the British Columbia Persons With AIDS Society), “Anti-Marijuana” campaign led by the Government of Canada, and “Make Health Last” campaign led by the Heart and Stroke Foundation. Because these campaigns differ in purpose, duration, and delivery channels, this may help identify components of a successful campaign that lead to increased information seeking behaviors. [Table 1](#) provides a summary and comparison of the campaign features.

### Data Collection

#### *Search Activity Given by Google Trends*

In 2012, Google Search accounted for 78% of the global market share among all search engines [30]. Aptly, its huge popularity will provide the best portrayal of online search activity in the Canadian population. Google Trends is a publicly available, free analytical tool that provides aggregated search results since 2004. The disease name was used as the base of each search term (ie, “colorectal cancer,” “HIV/AIDS,” “marijuana use,” and “stroke”). Furthermore, to produce an accurate portrayal of the search activity behaviors of the population, the related key terms suggested by Google Trends (see [Multimedia Appendix 1](#)) were reviewed for inclusion in the search. These suggested terms helped to capture the most popular search queries as well as synonymous terms. However, terms that were ambiguous, carried multiple meanings, and/or related to disease statistics were excluded from the data collection. For example, recommended Google searches such as “Hiv/aids in Africa,” “2 stroke,” “heart and stroke lottery,” and “hiv prevention” did possess the target word in the query but contextually were not relevant to the study and were subsequently not collected in this process. Using Boolean operators, the search query contained the keywords relevant to the disease in order for Google Trends to generate the aggregate search patterns. The final search queries used in this study are provided in [Table 2](#). Google Trends returned results in RSV, which facilitates comparisons between terms. The RSV is defined as the quotient of the number of searches at a specific time over the total number of searches in a period of time, for a specific geographic location. Essentially, the RSV represents the popularity of the search term at a specific point in time. After narrowing filters in the Google Trends tool, weekly RSV data were exported for the period February 2004 to December 2015 in Canada and the specific Canadian provinces that were associated with the campaigns.

**Table 1.** Summary of health campaigns.

Campaign features	Disease interests			
	Colon cancer	HIV <sup>a</sup> and AIDS <sup>b</sup>	Marijuana use	Stroke
Organization	Government of Ontario	Positive Living Society of British Columbia	Government of Canada	Heart and Stroke Foundation
Campaign name	“ColonCancerCheck”	“End HIV Stigma”	“Anti-Marijuana”	“Make Health Last”
Purpose	Increase colon cancer screening practices [25]	Reduce stigma surrounding HIV [26]	Educate about the negative health consequences of marijuana among adolescents [27]	Increase awareness of stroke [28] Educate about risk assessment [28] Promote health [28] Encourage donations [28]
Delivery channels	Health care provider referrals as well as television advertisements, radio announcements, newspaper advertisements, and pamphlets across Ontario [25]	30-second public service announcements shown on 40 participating radio and television stations in British Columbia [26]	Television, Web-based advertisements, and social media [27]	Canadian Broadcasting Corporation (CBC) platforms including CBC television, CBC networks, CBC Player, regional stations, and digital banner [29]
Campaign period	April 2008 to September 2008 [25]	July 2006 to July 2007 [26]	October 2014 to December 2014 [27]	February 2013 to May 2013 [29]
Duration	6 months	12 months	3 months	4 months

<sup>a</sup>HIV: human immunodeficiency virus.

<sup>b</sup>AIDS: acquired immunodeficiency syndrome.

**Table 2.** List of search terms.

Search query filters	Colorectal cancer	HIV <sup>a</sup> /AIDS <sup>b</sup>	Marijuana use	Stroke
Search terms	Colorectal cancer + colorectal diagnosis + colorectal screening + colorectal cancer screening + colon cancer + colon cancer symptoms	Hiv + aids + human immunodeficiency virus + acquired immunodeficiency virus + hiv symptoms + hiv diagnosis + aids symptoms + aids diagnosis + hiv contraction	Marijuana use + drug abuse + marijuana side effects + marijuana effects + effects of marijuana + drug use + drug addiction	Stroke + stroke symptoms + stroke onset
Geographic locations studied	Canada and Ontario	Canada and British Columbia	Canada and Ontario	Canada and Ontario
Period of data collection	February 2004 to December 2015	February 2004 to December 2015	February 2004 to December 2015	February 2004 to December 2015

<sup>a</sup>HIV: human immunodeficiency virus.

<sup>b</sup>AIDS: acquired immunodeficiency syndrome.

### Disease Statistics

Statistics on the diseases were searched in both open peer-reviewed journals and gray literature sources. Data reported on the disease prevalence were preferred over disease incidence; however, if prevalence data were not available, then disease incidence data were still used. The disease incidence and prevalence data were obtained from Statistics Canada and other governmental publications for the period 2004 to 2014.

### Data Analysis

#### Relationship Between Search Activity and Disease Patterns

Data trends for each disease were graphed together to compare disease search activity and disease prevalence or incidence. Disease monitoring was first assessed via visual inspection to

identify patterns in the data. A subsequent Pearson correlation analysis was conducted in IBM SPSS Statistics 23 (IBM Corporation) to detect statistical significance. The mean annual search activity was plotted against the annual disease rates to determine any correlations. This procedure was repeated for all 4 diseases.

#### Impact of Health Campaign on Search Activity

To evaluate the public health campaign effect on search activity, the joinpoint methodology was used [11,31]. This analysis was used by Schootman et al [11] to investigate the impact of cancer diagnosis and deaths of public figures on search behaviors on Google. Similarly, in the time series of Google search activity, the joinpoint methodology tested whether weekly changes in RSV around the time of the campaign were statistically significant. This was completed by fitting linear regression lines to the natural logarithm of the weekly RSV. Inflection points

were identified to examine any significant differences between the slopes of the linear regression lines. Analyses were performed to examine online activities 20 weeks before the campaign, during the campaign, and 20 weeks after the campaign. Secondary analyses were conducted to examine shorter segments of 5, 10, and 15 weeks before, during, and after the campaign. The purpose of assessing different periods was to identify any potential effects of the health campaign on increasing people's search behavior on the Web. For example, examining the 5 weeks leading up to the campaign and the 5 weeks following the campaign detected any immediate change caused by the campaign implementation. However, because RSV data were collected on weekly measures, the number of observations tested by the joinpoint program was low, affecting the statistical power. Therefore, longer-term trends were investigated in 10-, 15-, and 20-week segments. A maximum of 3 joinpoints were fitted to each RSV time series. This maximum was chosen because of the predicted pattern we expected to see. The first joinpoint would be observed at the start of the campaign period because we hypothesized an increase in search activity. Following that surge, the search activities would peak during the campaign period and slightly decline over time, generating the second joinpoint. The third

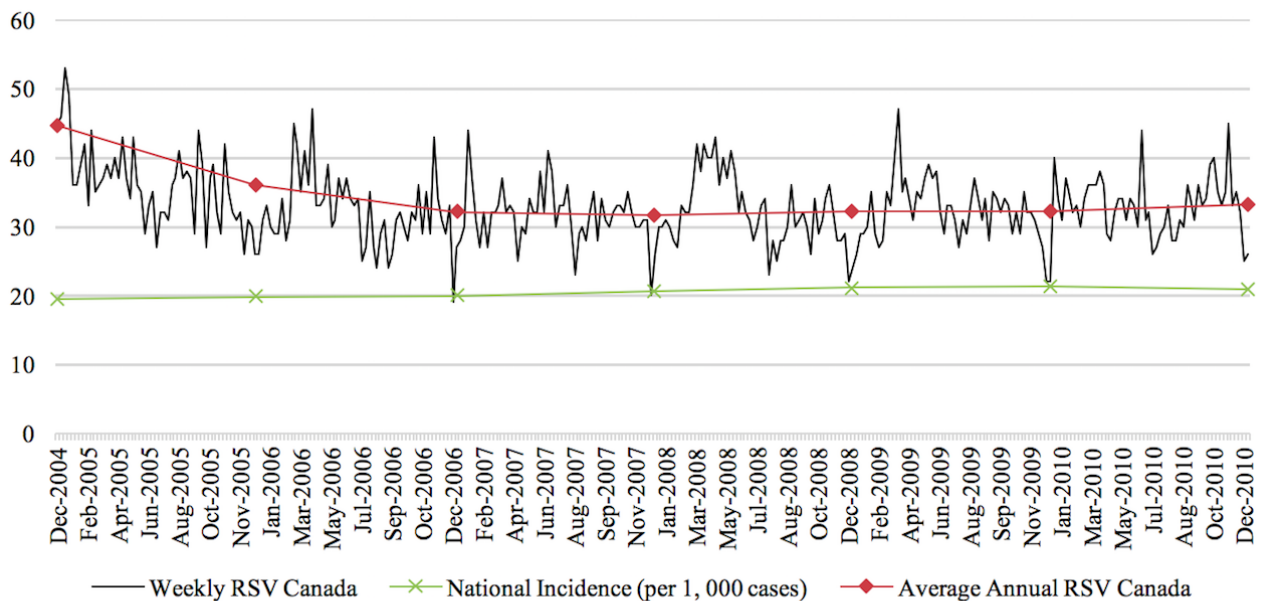
joinpoint would be observed at the end of the campaign period, in which search activity would return to baseline.

## Results

### Internet Search Activity and Disease Prevalence

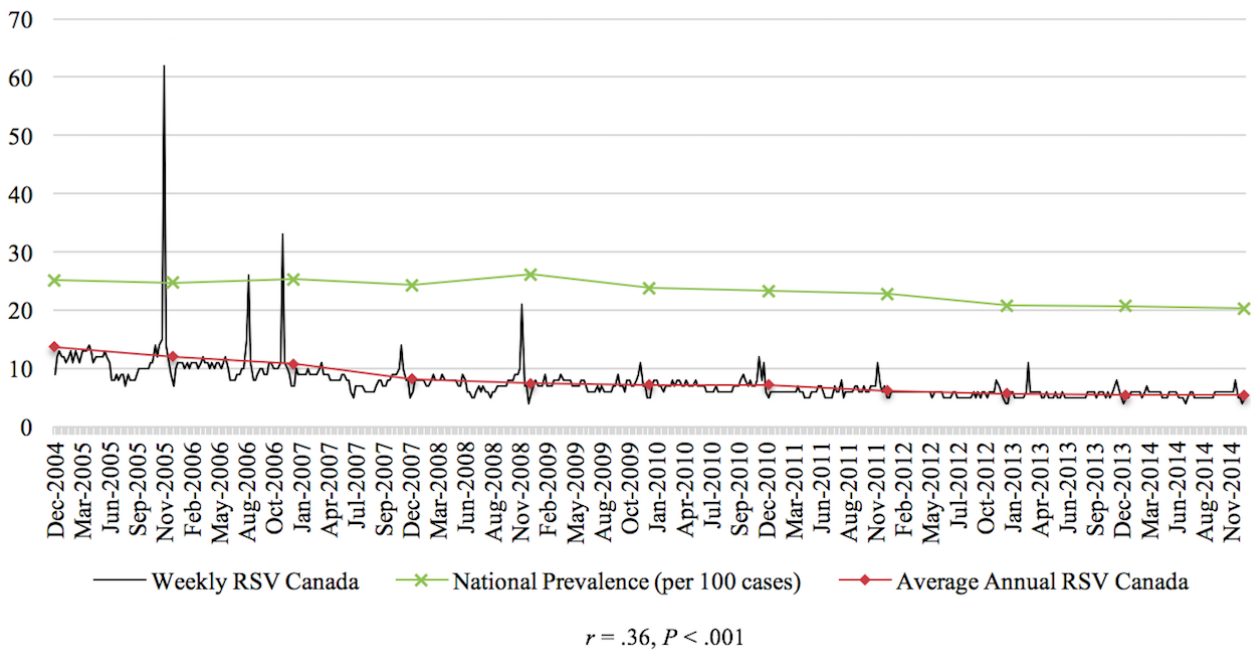
The Pearson correlations tested the relationship between RSV of the disease and the prevalence or incidence of the disease in a given time period for the 4 diseases studied in this paper. In general, the relationship between these 2 variables was significant for each of the 4 diseases. From 2004 to 2010, there was a negative correlation between colorectal cancer incidence and search activity ( $r = -.41, P < .001$ ). The association between HIV search activity and incidence data was moderately positive between 2004 and 2014 ( $r = .36, P < .001$ ). The prevalence of marijuana use pertaining to 2004 and from 2008 to 2012 was positively correlated with search activity ( $r = .64, P < .001$ ). Finally, from 2004 to 2009, the association between the search activity of stroke and the incidence was moderately positively correlated ( $r = .40, P < .001$ ). Figures 1-4 depict the Web-based search activity trend in weekly RSV values and the calculated average RSV values for the year compared with the disease rates for colorectal cancer, HIV infection, marijuana use, and stroke.

**Figure 1.** Web-based search activity and incidence trends for colorectal cancer. RSV: relative search volume.

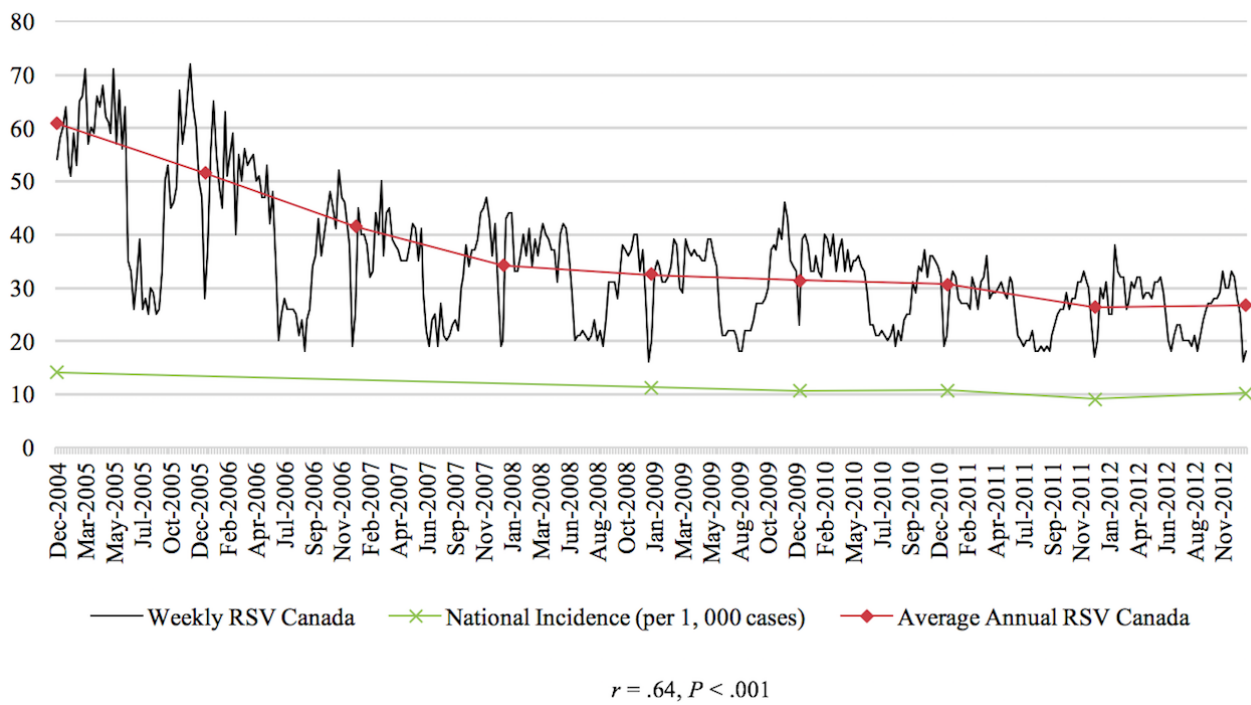


$r = -.41, P < .001$

**Figure 2.** Web-based search activity and prevalence trends for human immunodeficiency virus. RSV: relative search volume.

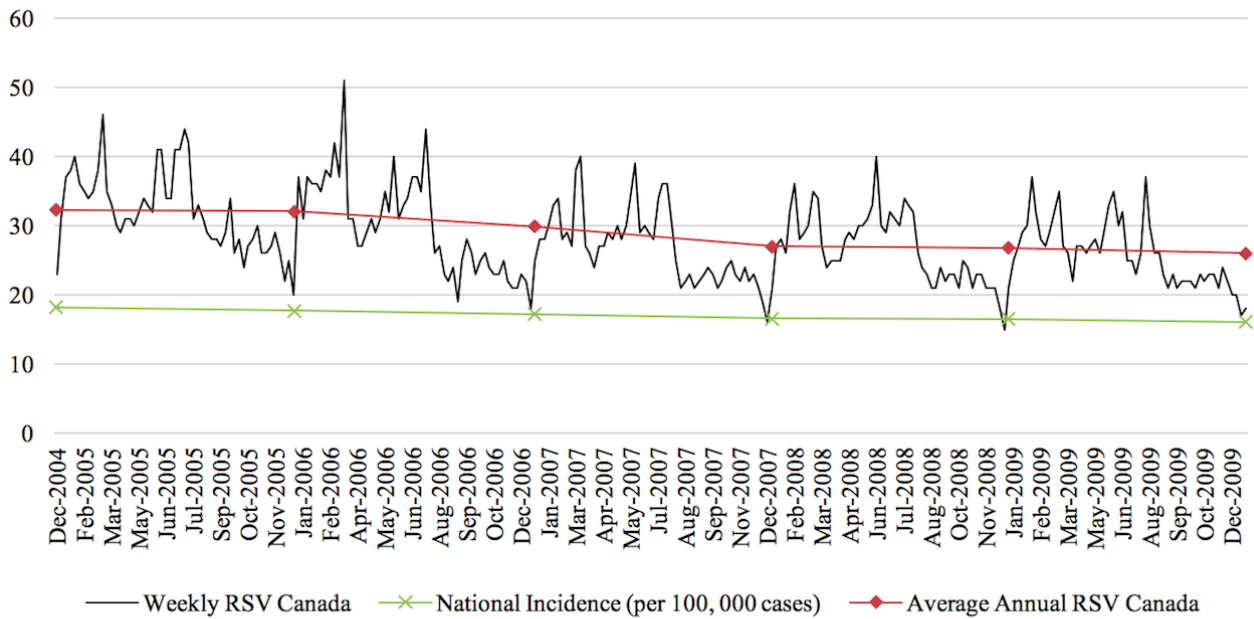


**Figure 3.** Web-based search activity and incidence trends for marijuana use. RSV: relative search volume.





**Figure 4.** Web-based search activity and incidence trends for stroke. RSV: relative search volume.



$r = .40, P < .001$

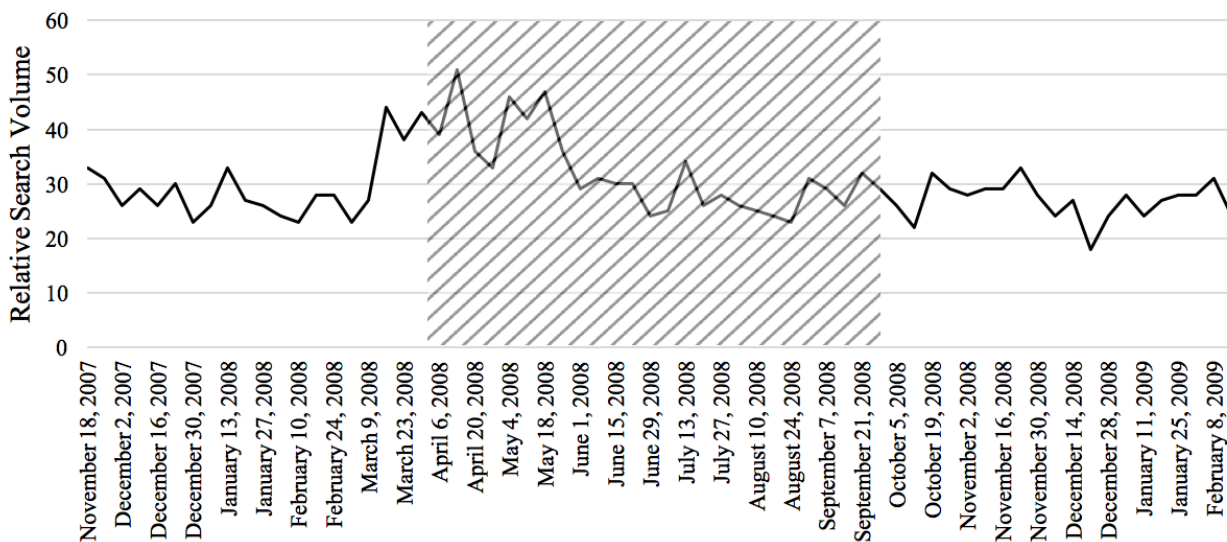
**Overview of Health Campaigns**

**Visual Assessment of Campaign Impact on Internet Search Activity**

The RSVs of the 20 weeks before the campaign, during the campaign period, and 20 weeks after the campaign were graphed to visually inspect them for increases and decreases in search activity caused by the implementation of the campaign

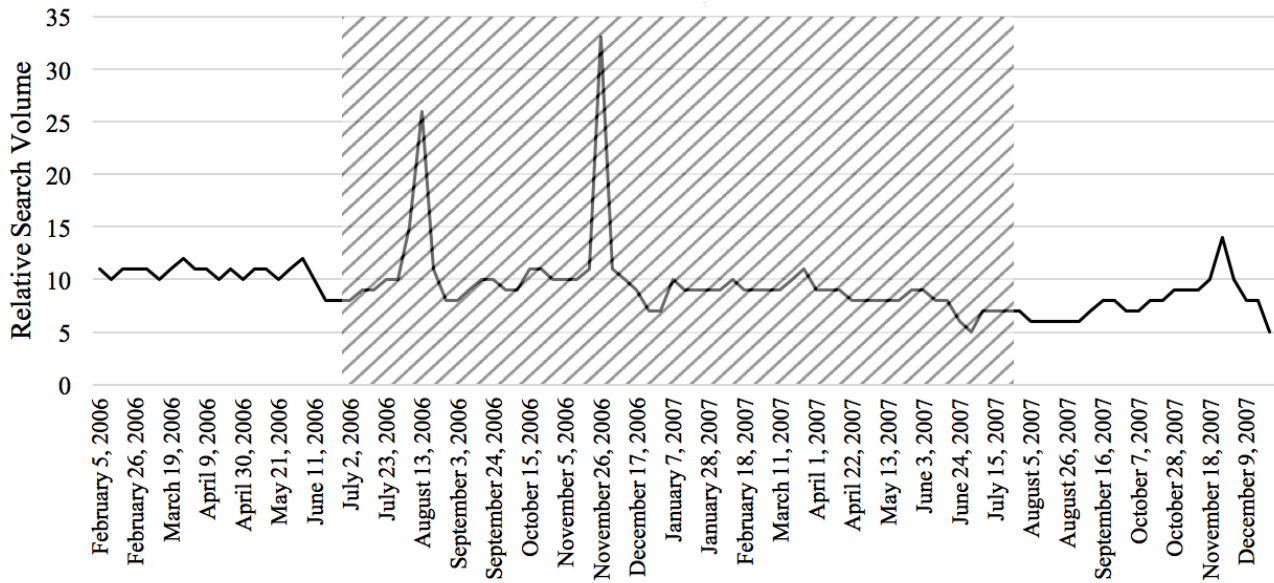
(examples shown in Figures 5-8). On the basis of this preliminary assessment, the “Anti-Marijuana” campaign (Figure 7) shows the largest increase in search activity during the campaign period than the other 3 campaigns. This is shown by the increase in activity as the campaign is introduced and the decrease in activity near the end of the campaign. No definitive inference could be made about the impact on information seeking behavior for the “Make Health Last” campaign for stroke and “End HIV Stigma” campaign.

**Figure 5.** Weekly Web-based search activity for colorectal cancer before, during, and after the campaign period. Highlighted section depicts the campaign duration.

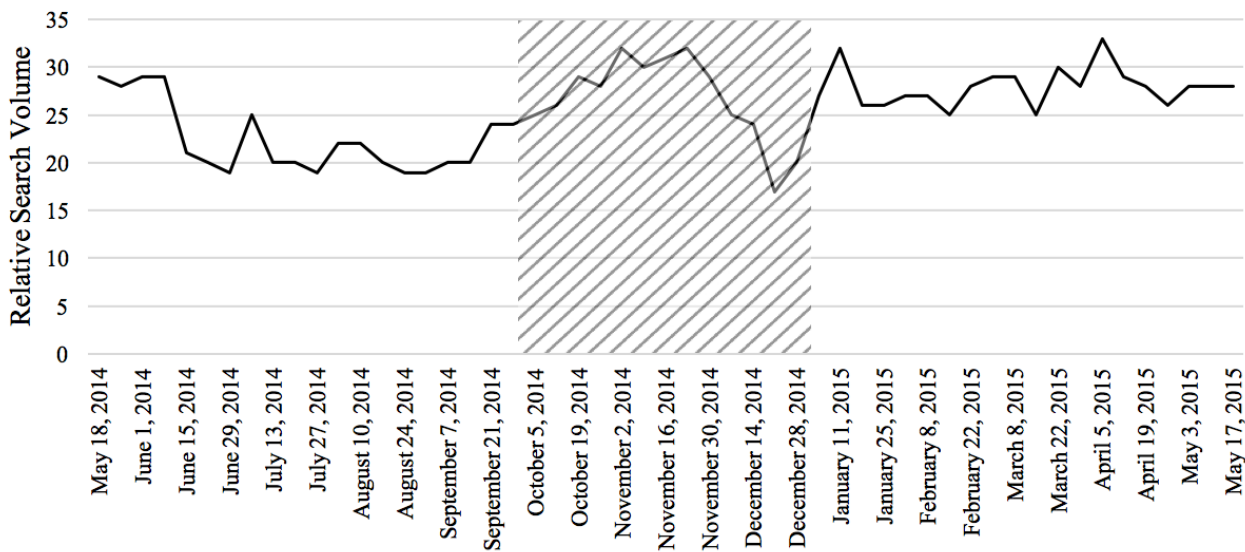


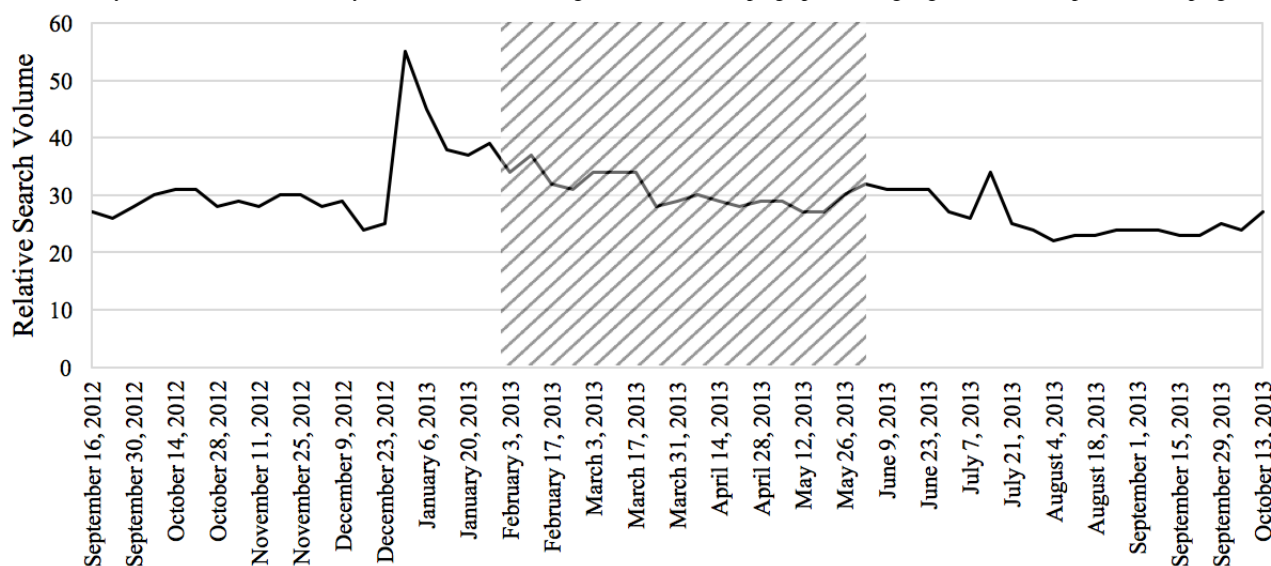


**Figure 6.** Weekly Web-based search activity for human immunodeficiency virus before, during, and after the campaign period. Highlighted section depicts the campaign duration.



**Figure 7.** Weekly Web-based search activity for marijuana use before, during, and after the campaign period. Highlighted section depicts the campaign duration.



**Figure 8.** Weekly Web-based search activity for stroke before, during, and after the campaign period. Highlighted section depicts the campaign duration.

### Statistical Findings of Campaign Impact on Internet Search Activity

Figure 9 shows the results from the joinpoint analysis. The results for “ColonCancerCheck” suggest strong effect of campaign implementation on Google search activity. A significant inflection point, which means a statistically significant change in slope, was detected at week 21. This point is very close to the 20-week mark of when the campaign began in our analysis. Notably, Figure 9 shows a declining trend of search activity between weeks 1 and 17 before experiencing a steep increase in search activity around the same time point of campaign implementation. The “Anti-Marijuana” campaign was the only campaign that showed an increase in search activity during the period of the campaign. The joinpoint analysis

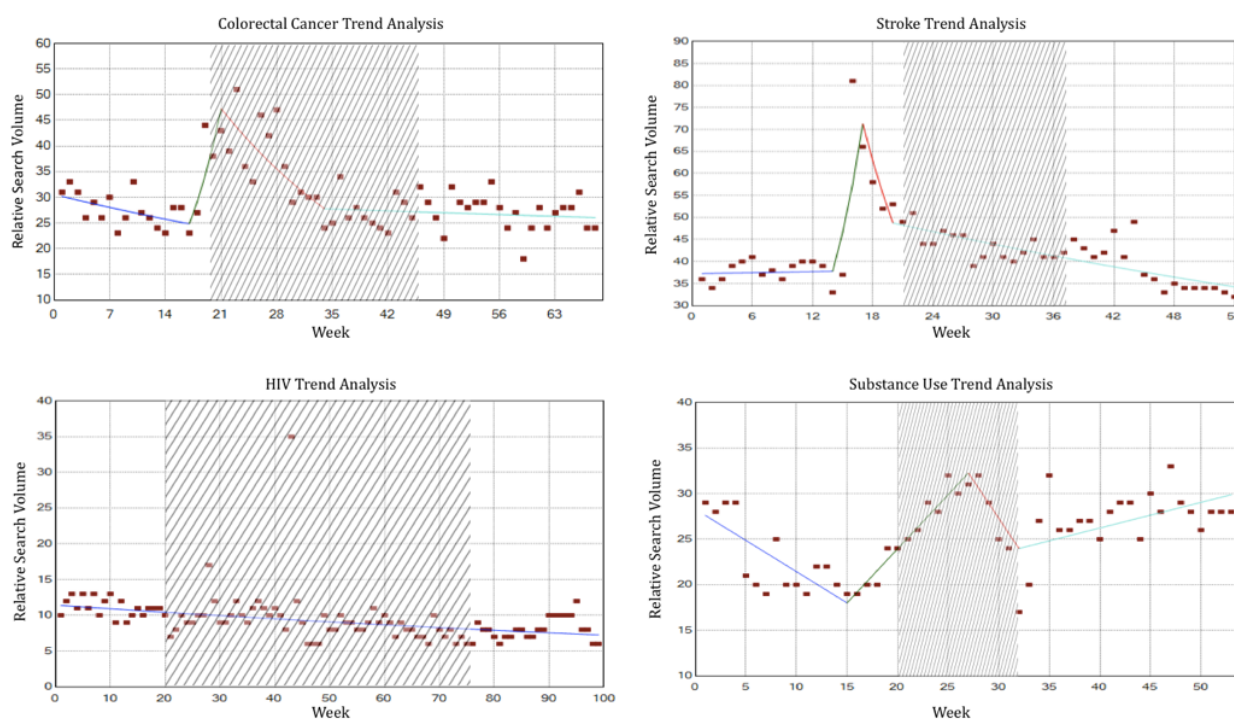
calculated a significant inflection point at week 16 in which search activity changes from a downward trend to an upward trend. The implementation of the campaign may have contributed to this observed increasing Google search activity for marijuana information. No significant associations between the campaigns and online search activity were seen for the “Make Health Last” and “End HIV Stigma” campaigns. The inflection points were not closely associated with the implementation of the campaign. Table 3 provides a summary of the statistical findings. A 20-week segment is provided as main results because it contains the most data points in its analyses; however, Multimedia Appendices 2-5 provide the statistical findings for the 5-, 10-, and 15-week period assessments.

**Table 3.** Joinpoint analysis for the periods 20 weeks before, during, and 20 weeks after the campaign.

Statistical outputs	Colorectal cancer	HIV <sup>a</sup>	Marijuana use	Stroke
<b>Segment 1 (week)</b>	1-17	1-99	1-16	1-14
Slope, RSV <sup>b</sup> /week (95% CI)	-1.21 (-2.7 to 0.3)	-0.46 (-0.6 to -0.3)	-2.72 (-4.1 to -1.3)	0.10 (-1.0 to 1.2)
P value <sup>c</sup>	.11	<.001	<.001	.85
<b>Segment 2 (week)</b>	17-21	—	16-28	14-17
Slope, RSV/week (95% CI)	17.46 (-2.8 to 41.9)	—	4.93 (2.7 to 7.3)	23.55 (-1.0 to 54.3)
P value <sup>c</sup>	.09	—	<.001	.06
<b>Segment 3 (week)</b>	21-34	—	28-33	17-20
Slope, RSV/week (95% CI)	-4.01 (-6.2 to -1.8)	—	-6.52 (-14.4 to 2.1)	-11.89 (-29.4 to 10.0)
P value <sup>c</sup>	<.001	—	.13	.26
<b>Segment 4 (week)</b>	34-68	—	33-55	20-54
Slope, RSV/week (95% CI)	-0.19 (-0.7 to 0.3)	—	1.19 (0.4 to 2.0)	-1.03 (-1.3 to 0.8)
P value <sup>c</sup>	.44	—	.006	<.001

<sup>a</sup>HIV: human immunodeficiency virus.<sup>b</sup>RSV: relative search volume.<sup>c</sup>Statistical significance was defined as  $P < .05$ .

**Figure 9.** Joinpoint analysis of the 4 diseases studied. The highlighted area shows the period the campaign was in effect. HIV: human immunodeficiency virus.



## Discussion

### Principal Findings

Both visual inspection and joinpoint analysis showed a positive correlation between health campaigns and people's search behaviors for the "ColonCancerCheck" and "Anti-Marijuana" campaigns. Although these associations were moderate, the results support previous studies, suggesting that infoveillance can measure the success of a campaign in driving information seeking behaviors in a population [10,32]. The "ColonCancerCheck" campaign was effective in generating online activity potentially because of the physicians being the point of contact to engage their target audience. As a result, this focused approach remained most relevant to at-risk populations, thus generating higher interest and information seeking behaviors. The impact of the "Anti-Marijuana" campaign on Google search activity may be due to the younger target audiences. The campaign aimed to educate adolescents about the negative effects of marijuana use on health. Fittingly, this target audience is also the primary user of the Web, which could have led to higher search activity for marijuana use information during the campaign period. However, no associations were found for the "Make Health Last" and "End HIV Stigma" campaigns. The null findings may be due to the differences in campaign type, frequency, and duration. Likewise, these same factors may understate the moderate associations observed for the "ColonCancerCheck" and "Anti-Marijuana" campaigns. Glynn et al [10] found positive associations between an annual breast cancer awareness campaign and related search activity in the United States. In their study, the authors observed a consistent increase in the level of Google search activity related to breast cancer during the month of October, which coincides with the annual breast cancer awareness month [10]. The success

of breast cancer awareness campaign in generating interest in the population, as measured by search activity on Google, may be easier to observe because of the recurrent pattern of the campaign. The campaigns examined in our study, on the other hand, were held only once for a specific period of time. Thus, the nonrepeating nature of the campaigns compared with the annual breast awareness campaign may have contributed to the little or no effect on Google search activity.

Another difference that may have implications on the findings is the length of the campaign period. Other studies examined more impulsive interventions and their immediate influence on online activity. Noar et al [23] looked at pancreatic cancer announcements from public figures and their influence on media and search query outcomes. From their study, they found a positive association between pancreatic cancer announcements, such as a diagnosis or death, and cancer information seeking behavior in the US population. In comparison, the durations of the campaigns in our study ranged from 3 months to 1 year. Consequently, the long durations of the campaigns may have weakened the observed effect of the campaign on generating interest and information seeking behaviors.

Despite past successes of linking Web-based search activity and infectious disease outbreaks, little is known about the relationship between Web-based search activity and chronic diseases. In this study, the disease rates of colorectal cancer, HIV infection, marijuana use, and stroke showed significant correlation with Google search activity. Although disease trends have not been studied yet in literature, these findings are consistent with other related studies. In one study, the state-specific variance in stroke prevalence was shown to be related to the search query data of the specific state [14]. Google search activity was also positively associated with suicide rates in the United States [18]. Stroke, HIV, and substance use showed

positive associations between disease rates and search volume. This means that more individuals with the disease would result in a corresponding increase in search activity observed on Google. Individuals may seek information on the Web before consulting a health care provider for reasons such as obtaining information about disease symptoms, diagnosis, or treatments [33]. Because of increasing Internet accessibility, more people are likely to engage in regular information seeking behaviors than in the past. Interestingly, colorectal cancer had a negative association with Web activity, which suggests that fewer individuals living with colorectal cancer would result in higher search activity on Google Trends. This inverse relationship between disease rates and online activity has not been reported in the literature. Although the reason behind this inverse relationship is not known, one possible explanation may be the decreased awareness of and interest toward colorectal cancer in Canada. In addition, cancer rates differ among the provinces and territories and an aggregate assessment may mask the true relationship between colorectal cancer prevalence and Internet search activity. Furthermore, due to the limited availability of information in scholarly journals and gray literature, colorectal cancer rates were only obtained from 2004 to 2010. It would be beneficial to assess the disease rates and search activity with more updated data to determine if the negative correlation remains persistent with time. Nonetheless, this relationship between monitoring cases of colorectal cancer and Web-based search activity should be further studied in order to confirm these findings.

### Limitations

Limitations of the study include the limited availability of information about the campaigns as well as disease statistics. Publicly available data on the prevalence of HIV, stroke, marijuana use, and colorectal cancer were limited. Ideally, disease prevalence data for the entire 2004 to 2015 period would provide the best Pearson correlation estimate. In this study, prevalence data were only available for HIV. Analyzing prevalence data would capture both individuals with newly diagnosed disease and individuals who are still currently living with the disease. Both groups should be considered because they are both potential users of the Internet who may seek more information about their disease. Compared with incidence data, this statistic only observes new cases of the disease over a period of time. Therefore, it neglects the second group of individuals who may still turn to the Web to seek health information about the disease. Consequently, individuals included in the incidence data would not be completely representative of those who are likely to use the Internet to learn about their disease.

A second limitation of this study is the primary use of Google Trends to collect and assess search activity. Although Google is currently the only search engine to offer a data analytics tool that is accessible to the public, there are biases present in using Google Trends. Because Google makes up 78% of the global market share [30], generalizability of the findings could still be maintained; however, this study still does not account for the remaining population who use other search engines such as Bing and Yahoo! [23]. Furthermore, crowdsourcing of data from the online community offers many advantages, such as readily available, easily accessible, and copious data to study.

Active crowdsourcing, such as reporting flu symptoms to a Web-based tracking system, or passive crowdsourcing, such as browsing the Web, would both come with inherent biases, which must be acknowledged [34]. This pertains to the volunteer bias in which individuals who use Google as their primary search engine were found to be typically younger, have higher income, and come from larger households [35]. These characteristics may act as confounders in this study as this specific population is less likely to develop the diseases, which in turn reduces their information seeking behaviors toward the diseases.

Finally, the general limitation in studying Internet search behaviors is the uncontrollable factors that can also affect search activity. Possible confounding variables include news events such as a disease-related death of a public figure, social media influence of a public figure, and other health campaigns held during the same time period. However, this is a classic limitation of conducting an observational study in which it would be impossible to discern a cause-effect relationship.

### Future Research

Further research is necessary to study associations between health campaigns and search activity as well as associations between disease rates and search activity on the Web. First, although the effect of health campaigns on Internet search activity was not established in this study, a recommendation for future studies is to examine the effects of recurrent health campaigns. Because positive findings have been reported for breast cancer awareness initiatives in the United States by Glynn et al [10], future studies could assess similar initiatives such as Movember for prostate cancer and their impact on search activity on the Web. Second, other analytics and statistical methods could be used to determine the effects of health campaign implementation and Internet search activity outcomes. Digital marketing techniques such as Buzzmonitor are used by organizations to track their brand popularity and “buzz” in the Web-based community [36,37]. Finally, big data are generated from many different sources that can also be studied to evaluate effectiveness of health campaigns. With the gaining popularity of social media platforms, such as Facebook, Twitter, and Instagram, among public health organizations, these data may offer potential answers to better understand the effects of health campaigns on the Web-based community. In a study by Xu et al [38], they evaluated health awareness campaigns using application programming interface to collect Twitter information. Their investigation showed fluctuations in the frequency of cancer terms on Twitter during the months of cancer awareness for breast, prostate, and lung cancers, namely, September, October, and November [38]. Similarly, these applications are transferable to health campaigns seeking related outcomes such as campaign awareness and health education. Thus, additional studies exploring new techniques and big data may help support the use of infoveillance in health campaign evaluation [36,37].

Furthermore, campaigns that focus more on increasing awareness may be more appropriate to study because information-seeking behavior is the target outcome of such campaigns. For example, campaigns using advertisements to deliver their message compared with those handing out



screening kits would be more relevant for engaging people on the Internet. Thus, selecting the right campaign will be an important factor to consider when studying search behaviors on Google Trends.

Application of Google Trends to provide indication of disease rates was shown to be promising in this study. However, future studies are still warranted in order to strengthen this correlation. First, disease prevalence rates should be examined instead of disease incidence rates because prevalent numbers would capture all individuals who are most likely to seek information regarding the disease on the Internet. In particular, diseases that have shown upward and downward trends in prevalence rates over a period of time will be best to study. With fluctuations in both positive and negative directions, it would test whether the search activity would also follow the fluctuations in both positive and negative directions. If both the disease rates and the Internet search activity agree within the same time frame, this correlation

would strengthen the use of infodemiology in monitoring diseases in a population.

## Conclusions

In this study, analysis of Internet search data showed significant relationships between health campaigns and information seeking behaviors in Canada for colorectal cancer and substance use but not for HIV and AIDS and stroke. The outcomes of the “ColonCancerCheck” and “Anti-Marijuana” campaigns were consistent with previous studies. Possible reasons for the discrepant findings from the “Make Health Last” and “End HIV Stigma” campaigns include differences in campaign type, frequency, and duration. However, the use of Web-based search data on digital disease monitoring remains promising. The study found significant associations between search activity and disease prevalence or incidence rates. Further studies are needed to validate the reliability of using Google Trends for health research purposes.

---

## Acknowledgments

The authors thank the undergraduate thesis program at the School of Public Health and Health Systems at the University of Waterloo for the opportunity to make this work possible.

---

## Conflicts of Interest

None declared.

---

## Multimedia Appendix 1

List of search terms reviewed for inclusion in the study.

[[PDF File \(Adobe PDF File\), 13KB - publichealth\\_v2i2e156\\_app1.pdf](#)]

---

## Multimedia Appendix 2

Results from joinpoint analysis for colorectal cancer.

[[PDF File \(Adobe PDF File\), 16KB - publichealth\\_v2i2e156\\_app2.pdf](#)]

---

## Multimedia Appendix 3

Results from joinpoint analysis for stroke.

[[PDF File \(Adobe PDF File\), 20KB - publichealth\\_v2i2e156\\_app3.pdf](#)]

---

## Multimedia Appendix 4

Results from joinpoint analysis for substance abuse.

[[PDF File \(Adobe PDF File\), 20KB - publichealth\\_v2i2e156\\_app4.pdf](#)]

---

## Multimedia Appendix 5

Results from joinpoint analysis for human immunodeficiency virus and AIDS.

[[PDF File \(Adobe PDF File\), 18KB - publichealth\\_v2i2e156\\_app5.pdf](#)]

---

## References

1. 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](https://doi.org/10.2196/jmir.1157)] [Medline: [19329408](https://pubmed.ncbi.nlm.nih.gov/19329408/)]



2. Eysenbach G. Infodemiology and infoveillance tracking online health information and cyberbehavior for public health. *Am J Prev Med* 2011 May;40(5 Suppl 2):S154-S158. [doi: [10.1016/j.amepre.2011.02.006](https://doi.org/10.1016/j.amepre.2011.02.006)] [Medline: [21521589](https://pubmed.ncbi.nlm.nih.gov/21521589/)]
3. Fernández-Luque L, Bau T. Health and social media: perfect storm of information. *Healthc Inform Res* 2015 Apr;21(2):67-73 [FREE Full text] [doi: [10.4258/hir.2015.21.2.67](https://doi.org/10.4258/hir.2015.21.2.67)] [Medline: [25995958](https://pubmed.ncbi.nlm.nih.gov/25995958/)]
4. Milinovich GJ, Avril SM, Clements AC, Brownstein JS, Tong S, Hu W. Using internet search queries for infectious disease surveillance: screening diseases for suitability. *BMC Infect Dis* 2014;14:690 [FREE Full text] [doi: [10.1186/s12879-014-0690-1](https://doi.org/10.1186/s12879-014-0690-1)] [Medline: [25551277](https://pubmed.ncbi.nlm.nih.gov/25551277/)]
5. Majumder MS, Santillana M, Mekaru SR, McGinnis DP, Khan K, Brownstein JS. Utilizing Nontraditional Data Sources for Near Real-Time Estimation of Transmission Dynamics During the 2015-2016 Colombian Zika Virus Disease Outbreak. *JMIR Public Health Surveill* 2016;2(1):e30 [FREE Full text] [doi: [10.2196/publichealth.5814](https://doi.org/10.2196/publichealth.5814)] [Medline: [27251981](https://pubmed.ncbi.nlm.nih.gov/27251981/)]
6. Nuti SV, Wayda B, Ranasinghe I, Wang S, Dreyer RP, Chen SI, et al. The use of google trends in health care research: a systematic review. *PLoS One* 2014;9(10):e109583 [FREE Full text] [doi: [10.1371/journal.pone.0109583](https://doi.org/10.1371/journal.pone.0109583)] [Medline: [25337815](https://pubmed.ncbi.nlm.nih.gov/25337815/)]
7. Brigo F, Trinka E. Google search behavior for status epilepticus. *Epilepsy Behav* 2015 Aug;49:146-149. [doi: [10.1016/j.yebeh.2015.02.029](https://doi.org/10.1016/j.yebeh.2015.02.029)] [Medline: [25873438](https://pubmed.ncbi.nlm.nih.gov/25873438/)]
8. Moore M, Loper KA, Van Schaik J. Seeking Medical Self-Diagnosis on the Web. Faculty Research, Publications, and Presentations Paper 29. Coral Gables, FL: University of Miami Scholarly Repository; 2012. URL: [http://scholarlyrepository.miami.edu/health\\_informatics\\_research/29/](http://scholarlyrepository.miami.edu/health_informatics_research/29/) [accessed 2016-10-01] [WebCite Cache ID 6kwmSSnTm]
9. Ayers JW, Althouse BM, Allem J, Rosenquist JN, Ford DE. Seasonality in seeking mental health information on Google. *Am J Prev Med* 2013 May;44(5):520-525. [doi: [10.1016/j.amepre.2013.01.012](https://doi.org/10.1016/j.amepre.2013.01.012)] [Medline: [23597817](https://pubmed.ncbi.nlm.nih.gov/23597817/)]
10. Glynn RW, Kelly JC, Coffey N, Sweeney KJ, Kerin MJ. The effect of breast cancer awareness month on internet search activity--a comparison with awareness campaigns for lung and prostate cancer. *BMC Cancer* 2011;11:442 [FREE Full text] [doi: [10.1186/1471-2407-11-442](https://doi.org/10.1186/1471-2407-11-442)] [Medline: [21993136](https://pubmed.ncbi.nlm.nih.gov/21993136/)]
11. Schootman M, Toor A, Cavazos-Rehg P, Jeffe DB, McQueen A, Eberth J, et al. The utility of Google Trends data to examine interest in cancer screening. *BMJ Open* 2015;5(6):e006678 [FREE Full text] [doi: [10.1136/bmjopen-2014-006678](https://doi.org/10.1136/bmjopen-2014-006678)] [Medline: [26056120](https://pubmed.ncbi.nlm.nih.gov/26056120/)]
12. Zhang Z, Zheng X, Zeng DD, Leischow SJ. Information seeking regarding tobacco and lung cancer: effects of seasonality. *PLoS One* 2015;10(3):e0117938 [FREE Full text] [doi: [10.1371/journal.pone.0117938](https://doi.org/10.1371/journal.pone.0117938)] [Medline: [25781020](https://pubmed.ncbi.nlm.nih.gov/25781020/)]
13. Abedi V, Mbaye M, Tsivgoulis G, Male S, Goyal N, Alexandrov AV, et al. Internet-based information-seeking behavior for transient ischemic attack. *Int J Stroke* 2015 Dec;10(8):1212-1216. [doi: [10.1111/ijss.12593](https://doi.org/10.1111/ijss.12593)] [Medline: [26282685](https://pubmed.ncbi.nlm.nih.gov/26282685/)]
14. Walcott BP, Nahed BV, Kahle KT, Redjal N, Coumans J. Determination of geographic variance in stroke prevalence using Internet search engine analytics. *Neurosurg Focus* 2011 Jun;30(6):E19. [doi: [10.3171/2011.2.FOCUS1124](https://doi.org/10.3171/2011.2.FOCUS1124)] [Medline: [21631220](https://pubmed.ncbi.nlm.nih.gov/21631220/)]
15. Willard SD, Nguyen MM. Internet search trends analysis tools can provide real-time data on kidney stone disease in the United States. *Urology* 2013 Jan;81(1):37-42. [doi: [10.1016/j.urology.2011.04.024](https://doi.org/10.1016/j.urology.2011.04.024)] [Medline: [21676450](https://pubmed.ncbi.nlm.nih.gov/21676450/)]
16. Wang H, Chen D, Yu H, Chen Y. Forecasting the Incidence of Dementia and Dementia-Related Outpatient Visits With Google Trends: Evidence From Taiwan. *J Med Internet Res* 2015;17(11):e264 [FREE Full text] [doi: [10.2196/jmir.4516](https://doi.org/10.2196/jmir.4516)] [Medline: [26586281](https://pubmed.ncbi.nlm.nih.gov/26586281/)]
17. Yang AC, Tsai S, Huang NE, Peng C. Association of Internet search trends with suicide death in Taipei City, Taiwan, 2004-2009. *J Affect Disord* 2011 Jul;132(1-2):179-184. [doi: [10.1016/j.jad.2011.01.019](https://doi.org/10.1016/j.jad.2011.01.019)] [Medline: [21371755](https://pubmed.ncbi.nlm.nih.gov/21371755/)]
18. Gunn JF, Lester D. Using google searches on the internet to monitor suicidal behavior. *J Affect Disord* 2013 Jun;148(2-3):411-412. [doi: [10.1016/j.jad.2012.11.004](https://doi.org/10.1016/j.jad.2012.11.004)] [Medline: [23182592](https://pubmed.ncbi.nlm.nih.gov/23182592/)]
19. Foroughi F, Lam A, Lim M, Saremi N, Ahmadvand A. "Googling" for Cancer: An Infodemiological Assessment of Online Search Interests in Australia, Canada, New Zealand, the United Kingdom, and the United States. *JMIR Cancer* 2016 May 04;2(1):e5. [doi: [10.2196/cancer.5212](https://doi.org/10.2196/cancer.5212)]
20. Jacobsen GD, Jacobsen KH. Health awareness campaigns and diagnosis rates: evidence from National Breast Cancer Awareness Month. *J Health Econ* 2011 Jan;30(1):55-61. [doi: [10.1016/j.jhealeco.2010.11.005](https://doi.org/10.1016/j.jhealeco.2010.11.005)] [Medline: [21185097](https://pubmed.ncbi.nlm.nih.gov/21185097/)]
21. Swenson W, Lindblom J. The effect of cancer awareness month campaigns on internet search volume and incidence of cancer diagnoses. In: *J Clin Oncol* 2013 May;13 (15 Suppl):e17512. Presented at: URL: [http://meeting.ascopubs.org/cgi/content/abstract/31/15\\_suppl/e17512](http://meeting.ascopubs.org/cgi/content/abstract/31/15_suppl/e17512)
22. Davis NF, Smyth LG, Flood HD. Detecting internet activity for erectile dysfunction using search engine query data in the Republic of Ireland. *BJU Int* 2012 Dec;110(11 Pt C):E939-E942 [FREE Full text] [doi: [10.1111/j.1464-410X.2012.11237.x](https://doi.org/10.1111/j.1464-410X.2012.11237.x)] [Medline: [22587462](https://pubmed.ncbi.nlm.nih.gov/22587462/)]
23. Noar SM, Ribisl KM, Althouse BM, Willoughby JF, Ayers JW. Using digital surveillance to examine the impact of public figure pancreatic cancer announcements on media and search query outcomes. *J Natl Cancer Inst Monogr* 2013 Dec;2013(47):188-194. [doi: [10.1093/jncimonographs/igt017](https://doi.org/10.1093/jncimonographs/igt017)] [Medline: [24395990](https://pubmed.ncbi.nlm.nih.gov/24395990/)]
24. Cooper CP, Mallon KP, Leadbetter S, Pollack LA, Peipins LA. Cancer Internet search activity on a major search engine, United States 2001-2003. *J Med Internet Res* 2005 Jul 1;7(3):e36 [FREE Full text] [doi: [10.2196/jmir.7.3.e36](https://doi.org/10.2196/jmir.7.3.e36)] [Medline: [15998627](https://pubmed.ncbi.nlm.nih.gov/15998627/)]

25. Rabeneck L, Timmouth JM, Paszat LF, Baxter NN, Marrett LD, Ruco A, et al. Ontario's ColonCancerCheck: results from Canada's first province-wide colorectal cancer screening program. *Cancer Epidemiol Biomarkers Prev* 2014 Mar;23(3):508-515 [FREE Full text] [doi: [10.1158/1055-9965.EPI-13-0956](https://doi.org/10.1158/1055-9965.EPI-13-0956)] [Medline: [24443406](https://pubmed.ncbi.nlm.nih.gov/24443406/)]
26. Davis M. 25 Years of AIDS in BC Marked With Hard-Hitting Provincial Ad Campaign. Vancouver, BC: British Columbia Persons With AIDS Society; 2006 Jun 29. URL: <https://positivelivingbc.org/wp-content/uploads/2015/03/bcpwa-release-end-stigma.pdf> [WebCite Cache ID 6kQdcGPmy]
27. Goodman LA. 2014. Government Anti-Drug Campaign Created Without Help from Medical Groups URL: <http://www.cbc.ca/news/politics/government-anti-drug-campaign-created-without-help-from-medical-groups-1.2806570> [accessed 2016-09-10] [WebCite Cache ID 6kQdlx4W0]
28. Make Health Last. Heart & Stroke Foundation. Heartandstroke. 2016. URL: [http://www.heartandstroke.mb.ca/site/c.lgLSIVOyGpF/b.8601065/k.6447/Make\\_Health\\_Last\\_PSA.s.htm](http://www.heartandstroke.mb.ca/site/c.lgLSIVOyGpF/b.8601065/k.6447/Make_Health_Last_PSA.s.htm) [accessed 2016-10-01] [WebCite Cache ID 6kw19f4Hk]
29. Richards E. CBC and Heart & Stroke Foundation Partners Again. *Marketingmag*. 2013. URL: <http://www.marketingmag.ca/brands/cbc-and-heart-stroke-foundation-partners-again-73018> [accessed 2016-09-13] [WebCite Cache ID 6kVAQZa14]
30. Simpson T. Evaluating Google as an Epistemic Tool. *Metaphilosophy* 2012 Jul 16;43(4):426-445. [doi: [10.1111/j.1467-9973.2012.01759.x](https://doi.org/10.1111/j.1467-9973.2012.01759.x)]
31. Surveillance.cancer. Joinpoint Trend Analysis Software URL: <https://surveillance.cancer.gov/joinpoint/> [accessed 2016-09-10] [WebCite Cache ID 6kQeAZjgg]
32. Bernardo TM, Rajic A, Young I, Robiadek K, Pham MT, Funk JA. Scoping review on search queries and social media for disease surveillance: a chronology of innovation. *J Med Internet Res* 2013;15(7):e147 [FREE Full text] [doi: [10.2196/jmir.2740](https://doi.org/10.2196/jmir.2740)] [Medline: [23896182](https://pubmed.ncbi.nlm.nih.gov/23896182/)]
33. Bader JL, Theofanos MF. Searching for cancer information on the internet: analyzing natural language search queries. *J Med Internet Res* 2003 Dec 11;5(4):e31 [FREE Full text] [doi: [10.2196/jmir.5.4.e31](https://doi.org/10.2196/jmir.5.4.e31)] [Medline: [14713659](https://pubmed.ncbi.nlm.nih.gov/14713659/)]
34. Swan M. Crowdsourced health research studies: an important emerging complement to clinical trials in the public health research ecosystem. *J Med Internet Res* 2012;14(2):e46 [FREE Full text] [doi: [10.2196/jmir.1988](https://doi.org/10.2196/jmir.1988)] [Medline: [22397809](https://pubmed.ncbi.nlm.nih.gov/22397809/)]
35. Freibott P. Who Uses Google, Yahoo! And Bing? Pmxagency. 2010 Aug 17. URL: <http://www.pmxagency.com/blog/2010/08/who-uses-google-yahoo-and-bing/> [accessed 2016-10-01] [WebCite Cache ID 6kwwkjqZf]
36. Getbuzzmonitor. GetBuzzmonitor: A Smarter Tool for Monitoring Social Media. 2016. URL: <http://www.getbuzzmonitor.com/> [accessed 2016-09-13] [WebCite Cache ID 6kV73n2H9]
37. Lima AB, Vitorino J, Rebêlo H. Cin. Buzzmonitor: a tool for measuring word of mouth level in on-line communities URL: [http://www.cin.ufpe.br/~hemr/papers/IADIS05\\_lima\\_vitorino\\_rebello.pdf](http://www.cin.ufpe.br/~hemr/papers/IADIS05_lima_vitorino_rebello.pdf) [accessed 2016-09-13] [WebCite Cache ID 6kVBRc9D]
38. Xu S, Markson C, Costello KL, Xing CY, Demissie K, Llanos AA. Leveraging Social Media to Promote Public Health Knowledge: Example of Cancer Awareness via Twitter. *JMIR Public Health Surveill* 2016;2(1):e17 [FREE Full text] [doi: [10.2196/publichealth.5205](https://doi.org/10.2196/publichealth.5205)] [Medline: [27227152](https://pubmed.ncbi.nlm.nih.gov/27227152/)]

## Abbreviations

**AIDS:** acquired immunodeficiency syndrome

**HIV:** human immunodeficiency virus

**RSV:** relative search volume

*Edited by G Eysenbach; submitted 16.08.16; peer-reviewed by O Leal Neto, D Jadhav; comments to author 31.08.16; revised version received 13.09.16; accepted 18.09.16; published 12.10.16.*

*Please cite as:*

Ling R, Lee J

*Disease Monitoring and Health Campaign Evaluation Using Google Search Activities for HIV and AIDS, Stroke, Colorectal Cancer, and Marijuana Use in Canada: A Retrospective Observational Study*

*JMIR Public Health Surveill* 2016;2(2):e156

URL: <http://publichealth.jmir.org/2016/2/e156/>

doi: [10.2196/publichealth.6504](https://doi.org/10.2196/publichealth.6504)

PMID: [27733330](https://pubmed.ncbi.nlm.nih.gov/27733330/)

©Rebecca Ling, Joon Lee. Originally published in *JMIR Public Health and Surveillance* (<http://publichealth.jmir.org>), 12.10.2016. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in *JMIR Public Health and Surveillance*, is properly cited. The complete bibliographic

information, a link to the original publication on <http://publichealth.jmir.org>, as well as this copyright and license information must be included.

Original Paper

# “Mommy Blogs” and the Vaccination Exemption Narrative: Results From A Machine-Learning Approach for Story Aggregation on Parenting Social Media Sites

Timothy R Tangherlini<sup>1</sup>, PhD; Vwani Roychowdhury<sup>2</sup>, PhD; Beth Glenn<sup>3</sup>, PhD; Catherine M Crespi<sup>3</sup>, PhD; Roja Bandari<sup>2</sup>, PhD; Akshay Wadia<sup>2</sup>, PhD; Misagh Falahi<sup>2</sup>, MS; Ehsan Ebrahimzadeh<sup>2</sup>, MS; Roshan Bastani<sup>3</sup>, PhD

<sup>1</sup>Center for Digital Humanities, University of California, Los Angeles, Los Angeles, CA, United States

<sup>2</sup>Department of Electrical Engineering, University of California, Los Angeles, Los Angeles, CA, United States

<sup>3</sup>Fielding School of Public Health, Jonsson Comprehensive Cancer Center, UCLA Kaiser Permanente Center for Health Equity, University of California, Los Angeles, Los Angeles, CA, United States

**Corresponding Author:**

Vwani Roychowdhury, PhD

Department of Electrical Engineering

University of California, Los Angeles

56-125B Engineering IV Building

420 Westwood Plaza (Box 951594)

Los Angeles, CA, 90095-1594

United States

Phone: 1 310 206 4975

Fax: 1 310 825 9754

Email: [vwani@ee.ucla.edu](mailto:vwani@ee.ucla.edu)

## Abstract

**Background:** Social media offer an unprecedented opportunity to explore how people talk about health care at a very large scale. Numerous studies have shown the importance of websites with user forums for people seeking information related to health. Parents turn to some of these sites, colloquially referred to as “mommy blogs,” to share concerns about children’s health care, including vaccination. Although substantial work has considered the role of social media, particularly Twitter, in discussions of vaccination and other health care–related issues, there has been little work on describing the underlying structure of these discussions and the role of persuasive storytelling, particularly on sites with no limits on post length. Understanding the role of persuasive storytelling at Internet scale provides useful insight into how people discuss vaccinations, including exemption-seeking behavior, which has been tied to a recent diminution of herd immunity in some communities.

**Objective:** To develop an automated and scalable machine-learning method for story aggregation on social media sites dedicated to discussions of parenting. We wanted to discover the aggregate narrative frameworks to which individuals, through their exchange of experiences and commentary, contribute over time in a particular topic domain. We also wanted to characterize temporal trends in these narrative frameworks on the sites over the study period.

**Methods:** To ensure that our data capture long-term discussions and not short-term reactions to recent events, we developed a dataset of 1.99 million posts contributed by 40,056 users and viewed 20.12 million times indexed from 2 parenting sites over a period of 105 months. Using probabilistic methods, we determined the topics of discussion on these parenting sites. We developed a generative statistical-mechanical narrative model to automatically extract the underlying stories and story fragments from millions of posts. We aggregated the stories into an overarching narrative framework graph. In our model, stories were represented as network graphs with actants as nodes and their various relationships as edges. We estimated the latent stories circulating on these sites by modeling the posts as a sampling of the hidden narrative framework graph. Temporal trends were examined based on monthly user-post statistics.

**Results:** We discovered that discussions of exemption from vaccination requirements are highly represented. We found a strong narrative framework related to exemption seeking and a culture of distrust of government and medical institutions. Various posts reinforced part of the narrative framework graph in which parents, medical professionals, and religious institutions emerged as key nodes, and exemption seeking emerged as an important edge. In the aggregate story, parents used religion or belief to acquire exemptions to protect their children from vaccines that are required by schools or government institutions, but (allegedly) cause

adverse reactions such as autism, pain, compromised immunity, and even death. Although parents joined and left the discussion forums over time, discussions and stories about exemptions were persistent and robust to these membership changes.

**Conclusions:** Analyzing parent forums about health care using an automated analytic approach, such as the one presented here, allows the detection of widespread narrative frameworks that structure and inform discussions. In most vaccination stories from the sites we analyzed, it is taken for granted that vaccines and not vaccine preventable diseases (VPDs) pose a threat to children. Because vaccines are seen as a threat, parents focus on sharing successful strategies for avoiding them, with exemption being the foremost among these strategies. When new parents join such sites, they may be exposed to this endemic narrative framework in the threads they read and to which they contribute, which may influence their health care decision making.

(*JMIR Public Health Surveill* 2016;2(2):e166) doi:[10.2196/publichealth.6586](https://doi.org/10.2196/publichealth.6586)

## KEYWORDS

vaccination; social media; machine learning; personal narratives; Internet; health knowledge; attitudes; practice

## Introduction

Over the past decade and a half, the explosion in social media and the concomitant rise in informational websites has changed the manner in which people access health care information [1-4]. Various sites dedicated to conversations about child rearing and parenting, colloquially referred to as “mommy blogs,” attract millions of users [3,5]. Although straightforward data mining techniques such as topic modeling exist for determining *what* parents are talking about on these sites and other similar sites, few techniques exist for determining *how* they are talking about those topics.

Among the many topics discussed on these parenting sites, few topics garner as much attention and vigorous discussion as childhood vaccination. Despite the fact that safe and effective vaccines exist, sporadic outbreaks of vaccine preventable diseases (VPDs) point to the continuing tension between public programs intended to make these vaccinations easily accessible and broadly adapted and parents who resist vaccination based largely on ideological principles [6-9]. Reduced rates of vaccination have jeopardized the elimination of diseases that have been on the cusp of such elimination for decades and, as recent outbreaks attest, threaten the hard-won herd immunity developed through long-term vaccination programs [6,10]. The role of exemptions in precipitating outbreaks in vaccine-communicable disease is increasingly being considered, although little evidence is currently available to directly support this link [11].

Although simple inspection of parenting sites and standard text mining approaches can confirm that vaccination is a topic of frequent discussion on these sites, such methods cannot determine the structure of those discussions. This is the objective of our research.

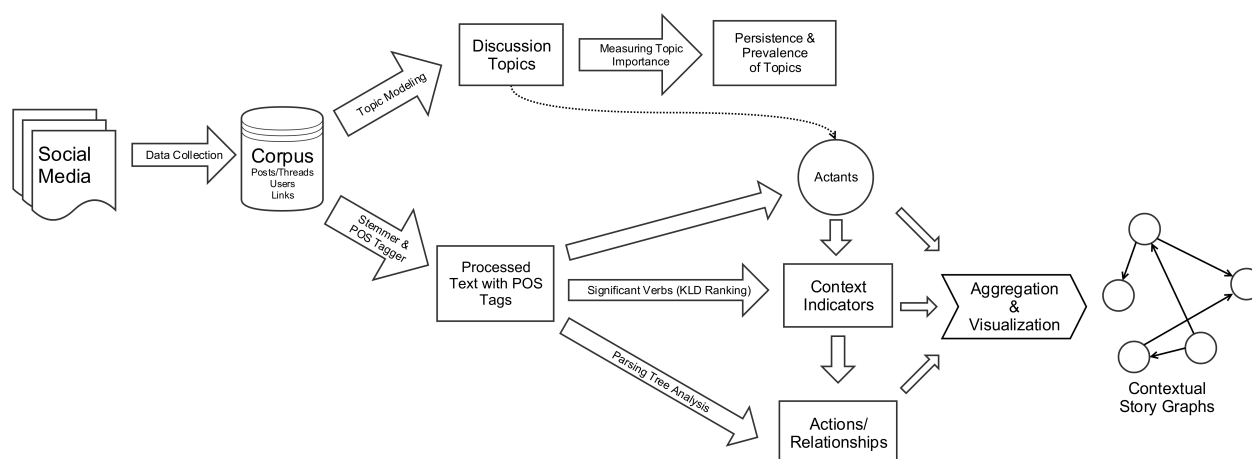
## Methods

### Introduction

In this research, we analyzed 1.99 million posts contributed by 40,056 users and viewed 20.12 million times indexed from 2 popular parenting sites over a period of 105 months ending in 2012. Beyond simply identifying the main topics of discussion on the sites, we discovered the underlying narrative frameworks that explain the stories circulating in these various discussions, an approach that extends recent work on personal experience and health knowledge exchange in Internet forums [12-16]. In addition to delineating the narrative framework that parents activate in their storytelling, we provided a fine-grained view of actant interactions and relationships in these stories, offering insight into individuals’ shifting attitudes toward vaccination. [Figure 1](#) shows the pipeline describing the steps of this workflow.

Data for this study were obtained from 2 popular social media sites dedicated to parenting. We chose these 2 sites because of their popularity among new parents, with a membership comprised primarily of people who self-identify as mothers [2,17]. As mothers are on the “frontline” of discussions about the health of their infant children, these sites offer important information about how they approach decisions related to vaccination [18]. Although the second site has a more ideologically diverse group of active posters than the historically anti-vaccination [mothering.com](#), both draw members from a wide range of backgrounds with broad geographic diversity, although largely from the United States and Canada. The language of both blogs is English. We indexed posts that appeared in forums related to childhood vaccination, recursively visiting and storing all publicly available discussion threads, and date-time–data, while creating an anonymized index of any accessible user data, resulting in a corpus of 299,778 posts from 12,376 users on [mothering.com](#), based on 105 months of indexed data (2004-2012) and 1,700,086 posts from 27,790 users on a second site (unnamed due to terms of service), based on 60 months of indexed data (2008-2012) (UCLA IRB #16-000456). These posts comprised the corpus for analyses.



**Figure 1.** Workflow for aggregate narrative framework discovery.

## Story Topics

For the 2 parenting sites, we determined the topics of discussion and the stories circulating in those discussions through an automated content analysis process. We started by computing dominant topics in the forums using 2 different probabilistic approaches, Latent Dirichlet Allocation (LDA) and Contextual Random Walk Traps (CRWT) [19,20]. LDA, a generative probabilistic topic modeling algorithm, proposes that each document in a corpus comprised consists of a mixture of a small number of topics, and that the appearance of a word in a document can be attributed to its membership in one of the document's topics. CRWT creates a cooccurrence network for all words in the corpus and then decomposes this network into a hierarchy of random-walk traps. Each such random-walk trap, comprising a series of document vocabularies (bag of words) weighted by their respective steady-state probabilities, serves as a topic in the CRWT. In both cases, we considered each thread (comprising a set of posts by different users) as 1 document. We use topic modeling solely to discover topics of discussion on these websites and derive ranked lists of nouns.

## Stories

To understand how people talked about the discovered topics, we developed a story model, the actant-relationship context model, and used it to extract the underlying stories from posts across the entire set of 1.99 million discussion posts, recognizing that forum posts frequently include only parts of stories or comments on story parts as opposed to complete stories. We conceptualize story parts as relationships among actants [21]. These relationships map well onto the vaccine story motifs described by Kitta [22] and make frequent use of the tropes discovered by Kata [3]. We developed a generative statistical-mechanical network model, in which actants (actors and objects) are the nodes, and the relationships between nodes are the edges. The edges are labeled with the nature of the relationship, the context in which the relationship was found, and its likelihood. Actants and relationships are then aligned with a modified version of Labov's 4-part structural map for personal experience narrative consisting of (1) orientation, where the community of interest is defined; (2) complicating action: threat, where something threatens the community delimited in

the orientation; (3) complicating action: strategy, where the actants in the story devise a strategy to deal with the threat; and (4) resolution, where the outcome of using the proposed strategy to thwart the threat is reported [23,24].

In our model, to generate a social media post, a user picks a set of actants and draws from the distribution of relationships among those actants. The user then composes the post according to the outcomes in the first step. In a social media corpus, the underlying probabilistic model including both the primary actants and their contextual relationships is hidden. Consequently, our task was to estimate this hidden model from the posts. We accomplished this through a computationally scalable estimation algorithm that requires minimal supervision. Because the data were large scale and the story signals were persistent, we found that a computationally scalable inference algorithm using minimal information (such as nouns and verbs) from Natural Language Processing (NLP) tools gave us accurate results for our dataset.

## Actants

We used the automatically discovered topics to determine the important actants in the topic space, recognizing that topics could cut across the siloes of forum classifications. To do so, we extracted a pool of actant terms based on a ranked list of high-frequency nouns. These nouns were, in turn, aggregated to derive actant categories. In topics associated with vaccination, we discovered 3 main categories of actants: individual actants, comprising parents, children, and medical professionals; institutional actants, comprising government institutions, religious institutions, pharmaceutical companies, and schools; and objects, comprising vaccines, exemptions, VPDs, and adverse effects. The words associated with an actant consist of both synonyms for the actant and entities that have the actant as a super-category. For example, the actant "government" includes the colloquial synonym "the Feds" as well as the government institution, the "CDC," where "government" is the super-category for CDC.

## Actant ↔ Actant Contexts

We characterized the context between a pair of actants by a set of verbs that are significant when the 2 actants are discussed simultaneously. Verbs are known to capture binary relationships

in large-scale corpora [25]. The contexts defined by verbs have discriminative power as they capture the different roles played by the same actants in different contexts.

In order to establish the significance of a verb for a particular pair of actants (ie, a context), we compared the conditional probability of the verb appearing with both actants to its marginal probability: A verb is *contextually significant* if  $P_{\text{pair}} = \text{Prob}(\text{verb} | \text{the sentence has both actants}) \gg P_{\text{corpus}} = \text{Prob}(\text{verb in any sentence in the corpus})$ . This approach attenuates the effect of commonly occurring verbs such as “has,” “is,” and “are” (for which  $P_{\text{pair}} \approx P_{\text{corpus}}$ ), while accentuating topical verbs that describe meaningful relationships between actants.

As there are many verbs involved in any context, we ranked the relative significance of the different verbs via a scoring or weighting function  $f(P_{\text{pair}}, P_{\text{corpus}})$ , and then selected the top ones as the verb set to characterize the context. We empirically tested various scoring functions, including term frequency-inverse document frequency (TF-IDF) style scoring functions, and discovered that the Kullback–Leibler (KL) divergence metric (Figure 2), produced the best results [26]. Whereas the results are largely invariant to the particular choice of the ranking method, we found that KL divergence was better

able to filter out noise such as the prevalence of modal and auxiliary verbs in the corpus. For any verb, the higher the KL score, the more significant that verb is to the pair.

To implement the above idea computationally, we tagged the entire corpus with parts of speech (POS) tags, using the Natural Language Toolkit (NLTK) library in Python [27]. As we extracted the verbs, we recorded their stemmed versions using the Porter stemmer in that toolkit. For example, “funded,” “funds,” and so on, were all recorded as the base form “fund.” For every stemmed verb,  $v$ , we calculated the marginal probability of the verb appearing in any sentence in the corpus (Figure 3), where  $N_v$  is the number of times verb  $v$  occurred in the corpus, and  $N$  is the sum of the frequencies of all the verbs in the corpus. Then, for any given context, defined as the set of all sentences where the 2 actants cooccur, we computed the conditional probability of a verb appearing with both actants in a particular context (Figure 4), where  $N_v(C)$  is the number of times verb  $v$  occurred in the given context, and  $N(C)$  is the sum of the frequencies of all the verbs in the context. Then we computed the ranking to determine the set of top verbs characterizing a given context (Figure 5) for all verbs and ranked them in decreasing order to obtain the set of top verbs that characterized the given context.

**Figure 2.** Calculation of the Kullback-Leibler divergence metric as a weighting function to rank the significance of different verbs.

$$f(P_{\text{pair}}, P_{\text{corpus}}) = \ln \frac{P_{\text{pair}}}{P_{\text{corpus}}}$$

**Figure 3.** Calculation of the marginal probability of the verb appearing in any sentence in the corpus.

$$P_{\text{corpus}}(v) = \frac{N_v}{N}$$

**Figure 4.** Calculation of the conditional probability of a verb appearing with both actants in a particular context.

$$P_{\text{pair}}(v) = \frac{N_v(C)}{N(C)}$$

**Figure 5.** Calculation of ranking to determine the set of top verbs characterizing a given context.

$$\ln \frac{P_{\text{pair}}(v)}{P_{\text{corpus}}(v)}$$

### Actant↔Actant Relationships

Once we had determined the ranked verb list for a context, we returned to the sentences for that context and determined the actant pairs that these significant verbs related using the POS tagger output. Recognizing that different verbs may capture the same type of relationship between actants, we grouped verbs into “relationship” categories, just as we grouped nouns into actant categories. Taking a cue from the narrative theory, we classified these relationships according to a series of binary oppositions between verbs, with highly ranked synonyms grouped together with their highly ranked antonyms, allowing us to align those relationships to the structure of the personal

experience narrative as well [28]. For readability, we devised labels for these groups of oppositional verbs.

We identified 2 main categories of binary opposite relationships. The first set of these relationships were those between individuals and institutional actants, with the binary oppositions *require or resist*, *advise or question*, *protect or threaten*, *employ or ignore*, *accept or reject*, and *attend or avoid*. The second set of relationships were those between individual and institutional actants on the one hand and objects on the other hand, or relationships between objects, with the binary oppositions *seek or aver*, *grant or withhold*, *cause or not cause*, and *protect or threaten*.

To illustrate this process, consider the verb “use” which was determined to be a significant verb in the Exemption $\leftrightarrow$ Religious-Institutions context (Table 1). In the following contextual sentence from a post that includes the actants *Exemption* and *Religious-Institutions*, we have highlighted the relevant words:

*Here is some New York info: (sample exemption letters here) Here is info about how you do not have to prove membership in a church in order to use a religious exemption*

The verb “use” relates “you” (which is a Parent actant) with “church” (which is a Religious Institutions actant). The category of the verbs that connects the 2 actants becomes the significant relationship between those actants in that context. For example, in the above case, as the verb “use” falls into the *employ or ignore* category, there is a directed *employ or ignore* edge from the Parents node to the Religious-Institutions node. We repeated this process for all possible contexts (we include additional examples of individual posts in the Multimedia Appendix 1).

### Story Graphs

We visualized each context as a network story graph, with the actants as the nodes, and the significant relationships as the edges connecting the actants, thereby capturing the rich structures of relationships among actants for any context. We then create a summary graph by aggregating the story graphs for each context into a single graph. We label this summary graph a narrative framework.

### Story Signal Trends

To characterize temporal trends in new posting activity that concerned vaccination exemptions and new user activity concerning exemptions, for each site, we calculated (1) the monthly proportion of new posts that included the word “exemption” and (2) the proportion of new users each month who committed a post with the word “exemption” in it. As users have access to old as well as new posts, in order to characterize the fraction of post content pertaining to exemptions that would be visible to users of the forums, we also calculated over the study period the monthly cumulative proportion of posts that

included the word “exemption.” We produced a log-linear plot of the distribution of user-activity duration (in days) for the mothering.com site using a bin width of 3 months.

## Results

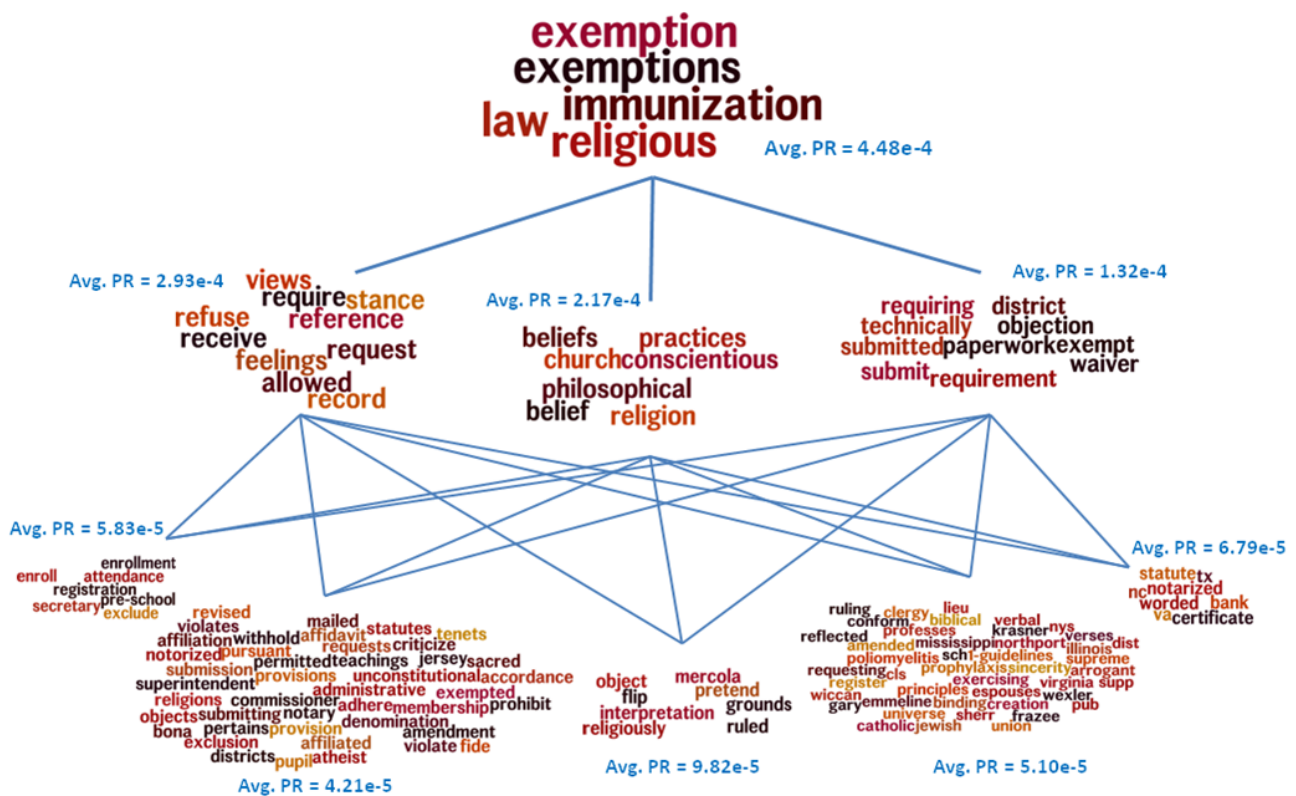
### Story Topics

On our 2 target sites, topic modeling revealed that vaccinations and, interestingly, exemptions constitute significant topics of discussion (for a full listing of topics, see the Multimedia Appendix 1).

We ran LDA topic modeling in R at multiple levels of granularity, from  $k=20$  to  $k=200$  in intervals of 20 (samples of the LDA topic models are included in the Multimedia Appendix 1) [29]. Past the topic parameter  $K=60$ , we found that topics, such as “exemption,” were largely split into multiple “exemption” topics. In the mothering.com forum, the “exemption” topic already emerged among the top topics with topic parameter,  $K=20$ : Topics 10 and 14 (Top 5 words): exemption school religious state required. Whereas the second site data presented a much larger set of forums and posts, the “exemption” topic nevertheless emerged in the top-60 list as a distinct topic (it did not appear as a distinct topic for  $K=20$  or 40, but rather as a significant part of larger topics such as vaccination): Topic 46 (Top 5 words): state religious exemption child form.

The CRWT method similarly yielded a more varied set of topics for the second site than for mothering.com, but the exemption-related topic was still distinct on both sites, constituted by some of the following words: “religious exemption beliefs exemptions belief belong supreme required.” As part of its output, CRWT yields a hierarchy of topics. For example, the “exemption” topic on mothering.com reveals a hierarchy where exemption is a super-category of “refusal,” “belief,” and “requirements,” as illustrated in Figure 6. These 2 topic modeling methods independently identify the importance of “exemption” as a topic of discussion in both forums at all levels of modeling that we used.

**Figure 6.** A hierarchical structure of topics related to exemption computed by the Contextual Random Walk Traps (CRWT) method from mothering.com posts. PR represents the page rank of the word-nodes in the co-occurrence network.



## Stories

The story model allows us to determine how people talk about the topics discovered through topic modeling. Recognizing that these topics can be discussed across the entire corpus, we do not assign documents to topics. Rather, we focus on discovering the underlying narrative framework, the activation of which in various posts, contributes to the structure of those discussions.

First, we determine the actants in the topic space (Table 1) and the relationship categories and associated verbs (Table 2). We find that the *Exemption*- and *Children*- related contexts are sufficient to derive a detailed understanding of the discussions and the stories that are embedded within them. The results for contexts based on exemptions (eg, exemptions and children) are shown in Table 3. The first column consists of the second actant in the relation, and the second column consists of the most significant verbs that occur in the related context. The stories clearly refer to families (orientation), with parents making health care decisions for their children. The verbs in the Parents↔Exemption context reveal that parents try to acquire exemptions as a strategy (complicating action: strategy). The motivation for seeking exemptions becomes apparent when one examines the verbs in the Children↔Vaccinations context (Table 4): seeking exemptions is a strategy to protect children from the (perceived) threat of vaccination (complicating action: threat). Results of the strategic use of exemptions vary from successfully securing exemptions to lamenting the inability to receive such exemptions.

We illustrate these findings with story graphs for different contexts, aggregating these into a single narrative framework graph (Figure 7). Importantly, our approach recognizes that the relationships among actants can vary depending on the context. In the Children↔Exemption story graph (Figure 7 a), one sees the central importance of exemptions for protecting children from required vaccines. The Exemption↔Schools story graph (Figure 7 b) shows the role that schools and other government institutions play in requiring vaccines (dark green), establishing an adversarial relationship between parents on one hand, and government institutions on the other. In the Exemption↔Religious Institutions story graph (Figure 7 c), an important relationship emerges between Religious Institutions and Parents (light blue), as we discover that parents and religious institutions are primarily connected on the basis of an *employ or ignore* relationship. This implies that parents use their affiliation with religious institutions (or the broader concept of faith) as a means to secure exemptions for their children. The contexts related to Children (Figure 7 d-e) further highlight the role that parents play in protecting their children from the adverse reactions that are allegedly caused by vaccination, as well as the role that affiliations to religious institutions can play in acquiring exemptions. In the Children↔Religious-Institutions context, for instance, the relationship between parents and religious institutions changes to one of accept or reject, focused primarily on parents' acceptance of church teachings.

**Table 1.** Actant model.

Entities (nodes)	Associated word set
<b>Individual actants</b>	
Parents	parents, parent, i, we, us, you
Children	child, kid, kids, children, daughter, daughters, son, sons, toddler, toddlers, kiddo, boy, d(ear)d(aughter), d(ear)s(on)
Medical professionals	doctor, doctors, pediatrician, pediatricians, nurse, nurses, ped, md, dr
<b>Institutional actants</b>	
Government	government, cdc, federal, feds, center for disease control, officials, politician, official, law
Religious institutions	faith, religion, pastor, pastors, parish, parishes, church, churches, congregation, congregations, clergy
Schools	teacher, teachers, preschools, preschool, school, schools, class, daycare, daycares, classes
Pharmaceutical companies	pharma, big pharma, company, companies
<b>Objects</b>	
Vaccines	vaccines, vax, vaccine, vaccination, vaccinations, shots, shot, vaxed, unvax, unvaxed, nonvaxed, vaccinate, vaccinated, vaxes, vaxing, vaccinating, substances, ingredients
Exemptions	exemption, exempt
VPDs <sup>a</sup>	varicella, chickenpox, flu, whooping cough, tetanus, pertussis, hepatitis, polio, mumps, measles, diphtheria
Adverse effects	autism, autistic, fever, fevers, reaction, reactions, infection, infections, inflammation, inflammations, pain, pains, bleeding, bruising, diarrhea, diarrhea

<sup>a</sup>VPDs: vaccine preventable diseases.

**Table 2.** Relationship model.

Relationships (edges)	Associated word set (stemmed)
<b>Between individuals or institutional actants</b>	
Require or resist	force, require, need, follow, mandate
Advise or question	recommend, tell, said, object, ask, learn, teach
Protect or threaten	protect, injure, damage
Employ or ignore	use, submit, ignore
Accept or reject	vaccinate, unvaccinate, vax, unvax, receive, have, had, get, inject, exclude, allow, exempt, believe, receive, request, deny, accept
Attend or avoid	enter, enroll, attend, go, send, homeschool
<b>Between individual or institutional actants and objects or between objects</b>	
Seek or aver	seek, file, sign, claim, submit, need, exercise, lie, claim
Grant or withhold	accept, approve, get, abuse, grant, oppose, deny
Protect or threaten	protect, injure, damage
Cause or not cause	expose, get, contract, cause, develop, suffer, die, vomit, diagnose



**Table 3.** Top 10 high-relevancy verbs (stemmed) that characterize the contexts comprising “Exemption” and each of the other major actant categories on the second site. The verbs are ordered according to the KL Divergence scores, but we have shown the frequency of the verbs in parenthesis for comparison. In the Exemption–Parent context, the verb “have” with a frequency count of 1561 is ranked fourth way before “exercise,” which has a frequency of only 275.

Actants in children context	Significant verbs in relationship to exemption (actant)
Parents	exempt(207), exercis(228), sign(275), have(1561), concern(196), claim(185), vaccin(241), belong(132), us(472), requir(199)
Children	exempt(220), exercis(191), concern(175), vaccin(202), vax(133), requir(152), sign(116), attend(99), enrol(56), allow(106)
Medical professionals	sign(101), exempt(21), give(72), write(34), requir(35), get(131), have(229), submit(16), obtain(17), file(17)
Government	bind(51), determin(51), requir(41), us(67), exempt(17), belong(22), accept(29), seek(19), furnish(8), obtain(12)
Religious institutions	belong(294), rule(114), offer(152), do(456), claim(105), us(191), have(445), find(149), bind(51), determin(50)
Schools	belong(146), rule(116), offer(160), requir(130), sign(120), attend(103), exempt(59), find(184), have(682), accept(104)
Vaccines	vaccin(377), exempt(173), requir(323), vax(252), claim(208), receiv(191), sign(150), request(106), allow(200), oppos(91)
VPDs <sup>a</sup>	requir(14), exempt(7), vaccin(15), sign(12), get(34), refus(9), have(55), prove(6), document(4), decid(8)
Adverse effects	had(64), exempt(12), obtain(12), requir(18), get(60), link(12), choose(9), follow(15), increas(11), qualifi(9)

<sup>a</sup>VPDs: vaccine preventable diseases.

**Table 4.** Top 10 high-relevancy verbs (stemmed) that characterize the contexts comprising “Children” and each of the other major actant categories on the second site. The verbs are ordered according to the KL Divergence scores, but we show the frequency of the verbs in parenthesis for comparison.

Actants in exemption context	Significant verbs in relationship to children (actant)
Parents	have(190359), give(27803), learn(18254), am(46173), choos(11446), want(46512), think(60272), rais(9756), know(61261), teach(8982)
Medical professionals	nurs(1728), vaccin(1450), told(1986), had(4091), said(2582), diagnos(893), take(2467), recommend(642), give(1782), took(926)
Government	vaccin(615), recommend(447), receiv(380), accord(291), mandat(144), ha(1211), caus(328), injur(139), report(214), includ(299)
Religious institutions	teach(2010), are(2873), is(3706), church(127), attend(221), go(937), rais(174), believ(335), allow(164), pray(73)
Schools	attend(2150), go(11284), ha(10885), send(2041), start(4659), work(4910), get(12171), daycar(621), need(5944), teach(1645), enrol(689)
Vaccines	vaccin(16262), vax(6176), receiv(3859), ha(8439), injur(1247), given(2321), unvaccin(930), caus(2315), recommend(1507), unvax(657), protect(1269)
VPDs <sup>a</sup>	vaccin(1409), receiv(1071), had(2622), get(2840), recommend(510), vax(450), given(590), develop(441), got(1005), expos(333)
Adverse effects	ha(17530), diagnos(4961), have(29879), had(9852), autism(1023), caus(2617), develop(1461), vaccin(1756), is(33778), affect(1064)

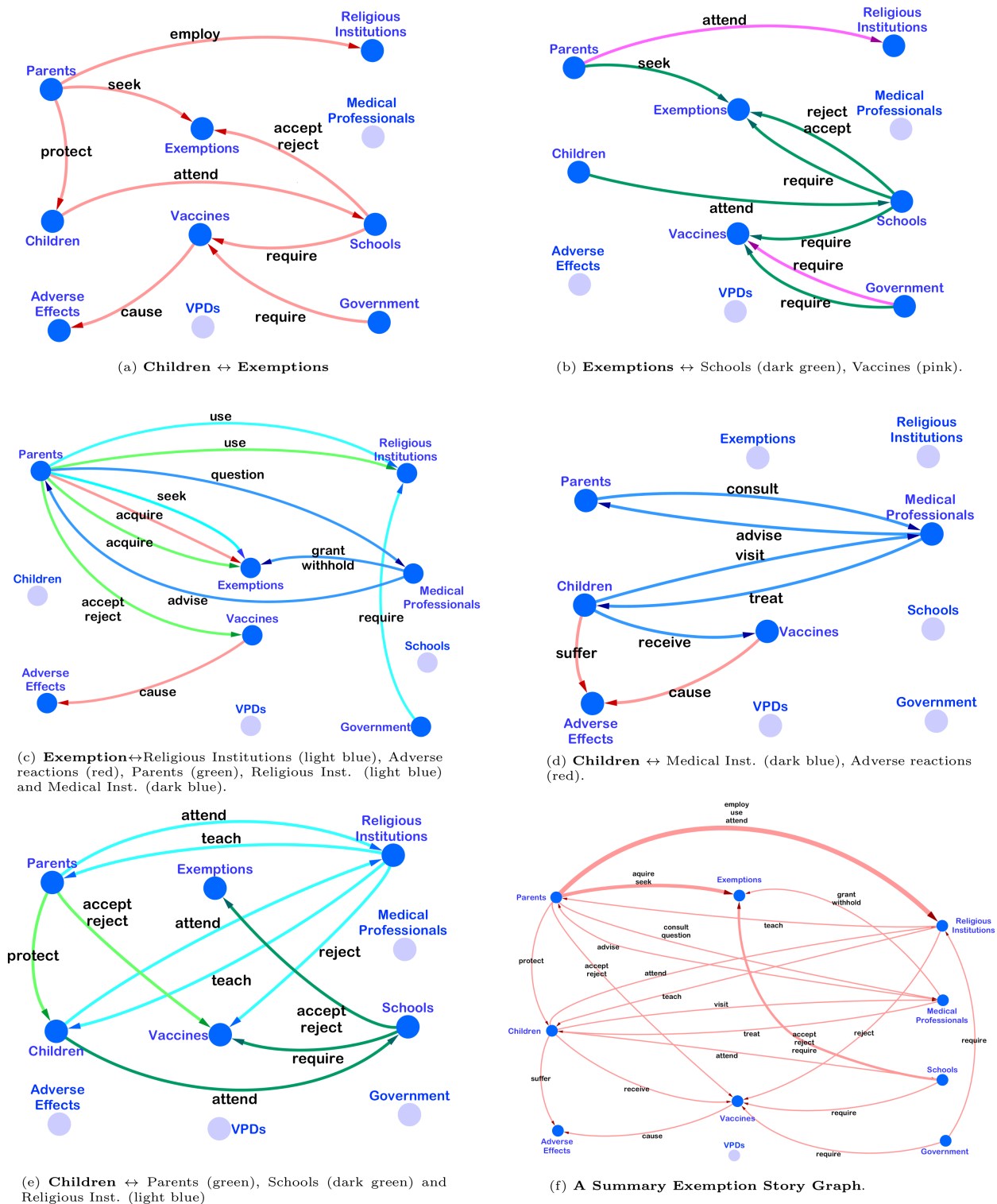
<sup>a</sup>VPDs: vaccine preventable diseases.

The summarized story graph (Figure 7 f), obtained by aggregating relationships across contexts, makes clear the underlying narrative framework. In the aggregate story, parents use religion or belief to acquire exemptions so as to protect their children from vaccines that are required by schools or government institutions, but (allegedly) cause adverse reactions such as autism, pain, compromised immunity, and even death.

The summarized story graph also reveals several notable substories. In one, religious institutions rather than schools play the role of “teacher.” In this substory, schools are relegated to the role of parental adversary, requiring vaccinations and wielding the power to accept or reject exemptions. In another

substory, medical professionals play the role of the adversary. Parents question them over the necessity of vaccines, and resent them as the enforcers of vaccine requirements (threat). Yet, parents also need the medical professionals’ help, as they act as the grantors of exemptions (strategy). Two glaring omissions in this and all the substories on which the summary narrative framework graph is based are the near total absence of VPDs and pharmaceutical companies as actants. The only role that the VPDs play is a passive one: children contract them (see the penultimate row in Table 2). “Big pharma,” as pharmaceutical companies are often referred to, play no significant role in the contexts in which exemptions are discussed.

**Figure 7.** Story graphs and narratives: subfigures a-e illustrate the story graphs corresponding to different contexts in mothering.com, while subfigure f is an aggregate master narrative graph.



**Story Signal Trends**

Figure 8 shows the proportion of all new posts that included the word “exemption” by month for each site. On average about 5.32% of new posts on mothering.com included the word “exemption”; on the second site, the average was about 0.35%. The trends show that the exemption topic signal exhibited some

variation over time but was not “bursty,” and there was some level of activity in all months, especially on mothering.com.

Figure 9 shows the proportion of new users of each site who committed a post containing the word “exemption” by month. On average, about 24.42% and 4.34% of new users committed such a post at the 2 sites, respectively. The proportions were fairly constant over the study period, with the exception of early

2007 for the second site, which was founded shortly before we began data collection.

Figure 10 shows the monthly cumulative proportion of posts that included the word “exemption” for the 2 sites. As users can view old as well as new posts, this metric helps characterize the pervasiveness of exemption-related content on the sites. On mothering.com, the fraction of posts pertaining to exemptions stabilized at a proportion of about 4.81%. For the second site, the cumulative fraction had an initial rise that then stabilized at about 0.34%.

A log-linear plot of the distribution of user-activity duration (in days) for the mothering.com site is presented in Figure 11. Note that the exponential tail (cut off) starts at around 2.5 years (ie, 1000 days), which is the typical age by which most children have received most of their vaccines. This suggests that new mothers are most active in the vaccination discussions and mothers become less active as their children pass the age when most vaccines have been administered.

Figure 8. Monthly proportion of new posts that included the work “exemption” for the two sites.

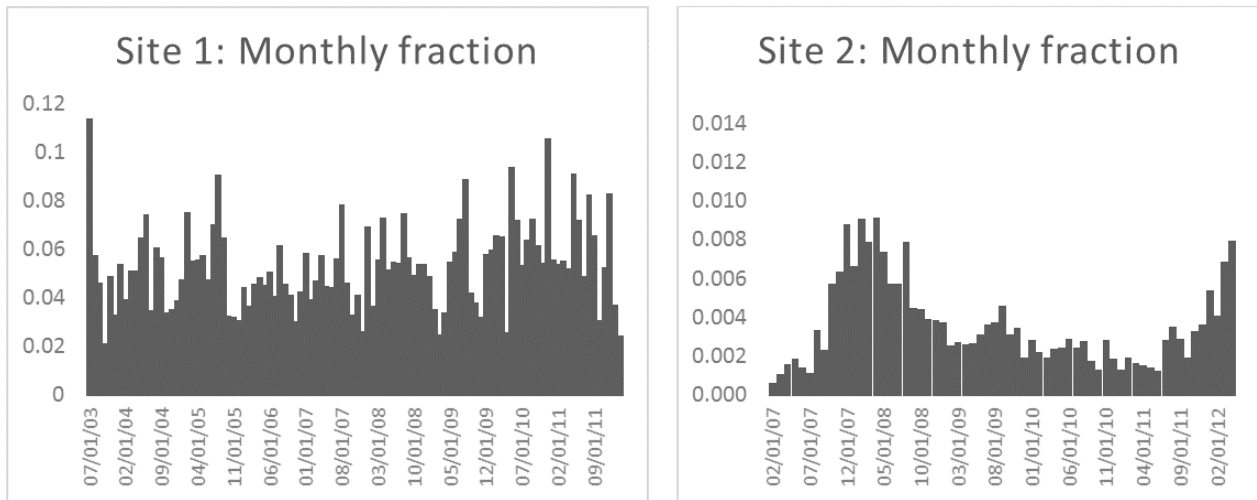
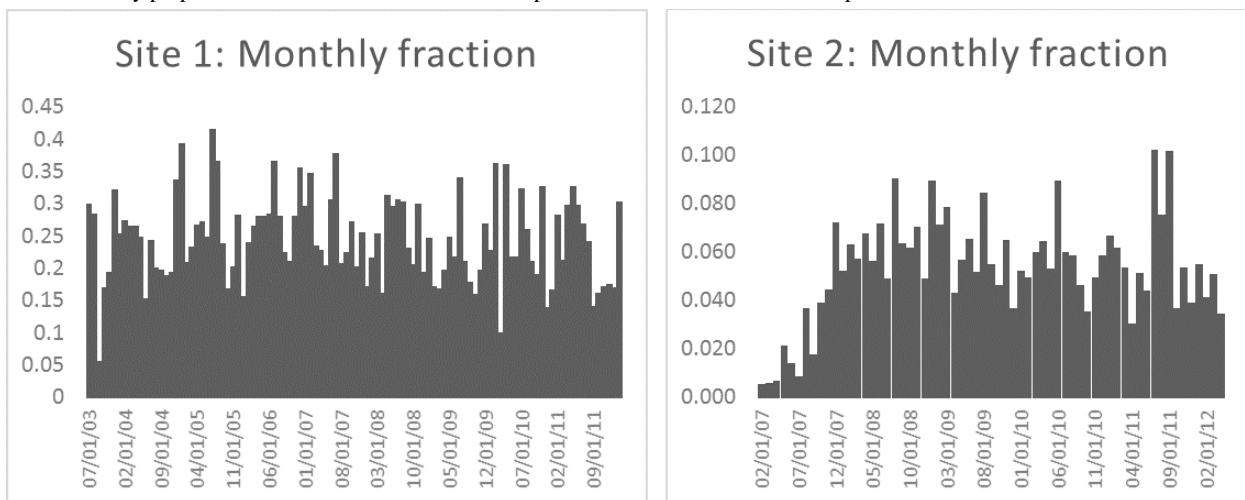
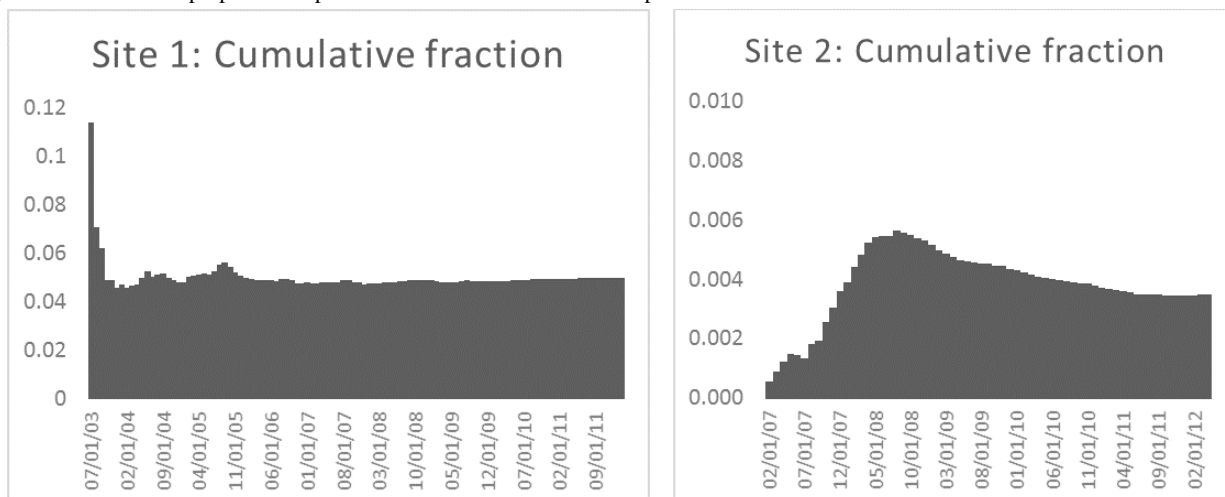


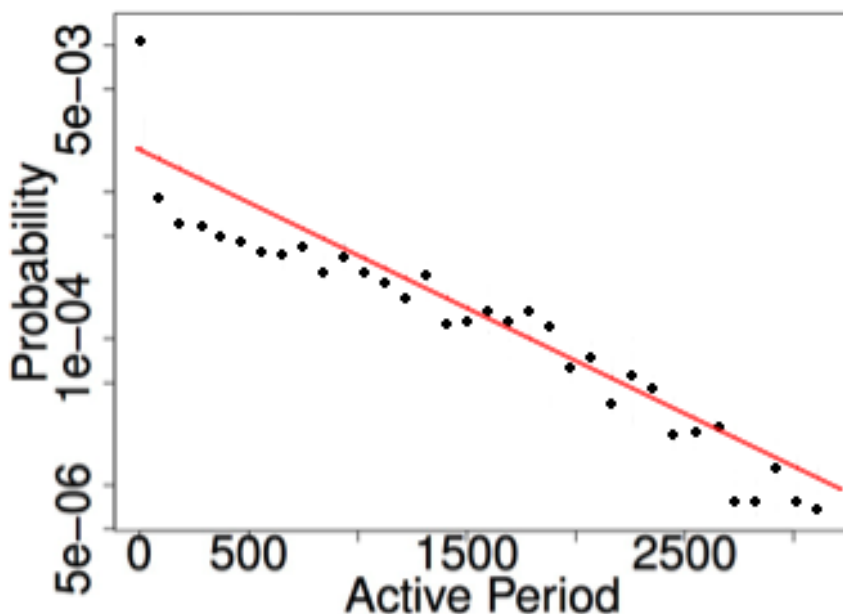
Figure 9. Monthly proportion of new users who committed a post that included the word “exemption” for the two sites.



**Figure 10.** Cumulative proportion of posts that contained the word “exemption” for each site.



**Figure 11.** Log-linear plot of the distribution of user-activity duration (in days) for the mothering.com site.



## Discussion

### Principal Findings

The methods we have developed for this study allow us to discover stories circulating informally on social media sites. Our system can detect the presence, persistence, and pervasiveness of story signals on otherwise very noisy sites, aggregate these story signals into a narrative framework, and provide a clear mechanism for tracing the emergence of specific strategies endorsed in these stories that parents might adopt to counteract perceived health-related threats. The sites play an important role in exposing parents to the ideas of vaccinations as threat, and the use of exemptions as a strategy to combat that threat [30]. Any new parent joining these sites, irrespective of their orientation to vaccination, is exposed to stories that activate the narrative framework of vaccination as threat and exemption as strategy.

On the sites we studied, the narrative framework is one where vaccines pose a threat to children, and parents in their role as protectors devise strategies, most often the use of exemptions, to thwart that threat. The narrative framework is so widely dispersed that it has traversed many segments of the parenting sites. A strong, persistent signal in these discussions reveals that parents actively pursued information about exemptions on an ongoing basis. New parents who joined these sites were likely to be quickly exposed to the beliefs encoded in these stories and the underlying narrative framework.

Given the well-established 90-9-1 rule of social media, where 90% of visitors simply read without commenting (9%) or contributing (1%) [31], it is very likely that the narrative framework is reaching a much larger audience than simple user statistics suggest. Note that the 1.99 million posts we studied had an aggregate view count of more than 20 million views from registered users (unregistered users could view the posts, but their views were not recorded and therefore not tabulated).

Even for parents who may not have initially believed that vaccines are harmful, the persistent circulation of stories about the potential harmfulness of vaccinations and the efficacy of the strategy of exemption to protect children from this alleged threat could convert some parents to embracing these beliefs [32].

### Limitations

Our work has certain limitations. As with all social media research, it is not clear that the sites on which we focused are representative of parents and health care decision makers as a whole. Although the 2 sites we used are very popular among parents, we recognize that they do not capture the broad range of discussions that take place in informal settings not easily observed, such as the playground, school gatherings, and other places where parents interact. In those settings, ethnographic fieldwork could provide important qualitative perspectives on vaccine-related discussions and storytelling [24]. At the same time, it is important to recognize that even settings that are more conducive to qualitative ethnographic methods are not immune to the influence of social media, which has increasing penetration into everyday life.

We recognize that parenting sites have certain biases. *Mothering.com*, for example, given its long-standing relationship with the now defunct *Mothering* magazine, has an anti-vaccination bias [11,12]. The well-known social network phenomenon of homophily might be creating an ideological echo chamber on these sites. Nevertheless, posts and user participation on the second site, which draws from a more ideologically diverse population, also reflect the persistent and prevalent nature of the vaccine-exemption narrative framework. Extending our approach to a broader sampling of parenting sites might mitigate the potential biases in these posts. At the same time, extending the approach to other types of social media where conversations are less organized, such as Facebook, might capture a broader series of narrative frameworks structuring vaccine-related conversations. Certain popular social media sites, such as Twitter, given the significant constraints on post length, are not included in this study.

Our work does not currently include sentiment detection. Although we advance the inclusion of social media in health care beyond topic discovery to an analysis of the underlying narrative structure of those discussions, we have not studied the manner in which those discussions are framed, which sentiment detection may be able to provide.

Data privacy continues to be a significant concern for social media research. In our study, we anonymized all of our data as part of the indexing process, and thus were unable to exploit certain features of individual user and user community data. The trade-offs between user privacy and research benefits are part of the constantly shifting terrain of social media research, and we chose to err on the side of privacy. Data access is becoming an equally significant problem, as social media corporations are greatly reducing access to data that people post and share on their sites. These limitations make it increasingly difficult to track large-scale conversations over long periods of time.

### Comparison With Prior Work

Vaccination decision-making has been broadly studied [33], and the impact of social media on health care decision-making is receiving increasing attention [34,35]. Several important studies have focused on people's social media reactions to emerging health care crises such as disease outbreaks, where news stories often drive participation [35,36]. The role of narrative on social media sites as a persuasive rhetorical strategy in regard to health care decision making has also been explored, although this important work is largely preliminary [3,13,22,37]. Our work shifts attention away from bursts of activity on sites such as Twitter to a complementary examination of long-term conversations that evolve over many months and even years, with a primary focus on the emerging and endemic narrative frameworks that inform these conversations, aligning with other large-scale studies of attitudes toward health care decisions [38-40].

Narrative is recognized as a key means for shaping belief. Radzikowski et al [13], in their study of reactions to the California measles outbreak, develop a narrative model based on keyword cooccurrence in Tweets, an excellent first pass at such a model, given the limiting factors that Twitter imposes on Tweet length. Given the length of the posts in our data, we are able to develop a more elaborate narrative model, allowing us to extend the pairwise associations among actants. Our method not only shows that these pairwise relations are the key part of the conversations, but also reveals how actants are related in a context-dependent manner. Grant et al [5] provide clear evidence for the impact of personal experience narrative on vaccination attitudes through the qualitative comparison of 4 websites. Kitta [22], who worked with a similar structural model for vaccine narrative, develops an important typology of vaccine stories, whereas Kata [3] determines the tropes that are functional on anti-vaccine websites. Our automated methods allow us to extend that qualitative work to very large-scale data (millions of posts), thereby operationalizing aspects of traditional text analytic methods.

Random sampling methods are another approach for understanding developing attitudes toward health care. However, the narratives we discover would be difficult to identify using random sampling approaches. Whereas the use of focus groups has shown itself to be particularly helpful in devising messaging campaigns for specific communities [41,42], understanding the narratives operative in a community and at the scale of big data may help refine those messages [43-45]. Mathematical modeling for vaccine decision-making [46], while promising, makes specific assumptions that need to be validated from real-world data for the results to be actionable. Our approach may help provide that data.

### Conclusions

Injecting an idea, such as the efficacy of exemption as a strategy to avoid vaccination, into online communities has the potential to influence many people—the idea can, in a phrase, “go viral.” Given the persuasive nature of personal experience narrative, storytelling plays a central role in exposing people to ideas and converting people to particular beliefs. Importantly, people are inclined to believe first-hand accounts from members of their



community, as opposed to official pronouncements [5]. Social network theory has established a strong tendency toward homophily in online communities that often results in shared trust between community members [47]. Notably, when “high degree” members (those with many connections in their network and to other networks) of a social network become exposed to beliefs and embrace them (in this case, the notion of exemptions), the conditions in the network become primed for rapid dissemination of those beliefs throughout the network. Unfortunately, once established, such beliefs are very difficult to change [33]. We believe that the personal stories highly popular on these sites make use of the shared trust developed in online forums and thus act as an ideal method for converting nodes to the beliefs encoded in those narratives. In our study, the persistence of the exemption signal suggests a broad-scale susceptibility in these networks to the exemption strategy for dealing with the “vaccination threat.”

The sheer volume of discussions on social media sites dedicated to parenting along with the knowledge that many people use Internet resources as their first line of health care information mean that these forums deserve ongoing attention [39,48]. Identifying the vaccine exemption narrative framework and its activation through individual storytelling is an important step in understanding how people discuss this topic on these sites. Similarly, identifying endemic signals, those with the greatest persistence and pervasiveness, can help separate ideas that have a very brief life span from underlying narrative frameworks that provide a foundation for repeated stories that contribute to ideas and attitudes becoming entrenched. Ultimately, our goal is to contribute to a system that monitors health care-related websites for emerging beliefs and attitudes, and that recognizes the power of narrative to persuade and create communities of like-minded individuals. Our work brings us a step closer to such a system.

---

## Acknowledgments

This work was supported in part by the NIH Grant Number R01 GM105033-01.

---

## Conflicts of Interest

None declared.

---

## Multimedia Appendix 1

Supplemental materials.

[PDF File (Adobe PDF File), 9MB - [publichealth\\_v2i2e166\\_app1.pdf](#) ]

---

## References

1. Zimmerman RK, Wolfe RM, Fox DE, Fox JR, Nowalk MP, Troy JA, et al. Vaccine criticism on the World Wide Web. *J Med Internet Res* 2005;7(2):e17 [FREE Full text] [doi: [10.2196/jmir.7.2.e17](#)] [Medline: [15998608](#)]
2. Kata A. A postmodern Pandora's box: anti-vaccination misinformation on the Internet. *Vaccine* 2010 Feb 17;28(7):1709-1716. [doi: [10.1016/j.vaccine.2009.12.022](#)] [Medline: [20045099](#)]
3. Kata A. Anti-vaccine activists, Web 2.0, and the postmodern paradigm—an overview of tactics and tropes used online by the anti-vaccination movement. *Vaccine* 2012 May 28;30(25):3778-3789. [doi: [10.1016/j.vaccine.2011.11.112](#)] [Medline: [22172504](#)]
4. Betsch C, Ulshöfer C, Renkewitz F, Betsch T. The influence of narrative v. statistical information on perceiving vaccination risks. *Med Decis Making* 2011;31(5):742-753. [doi: [10.1177/0272989X11400419](#)] [Medline: [21447730](#)]
5. Grant L, Hausman BL, Cashion M, Lucchesi N, Patel K, Roberts J. Vaccination persuasion online: a qualitative study of two provaccine and two vaccine-skeptical websites. *J Med Internet Res* 2015 May;17(5):e133 [FREE Full text] [doi: [10.2196/jmir.4153](#)] [Medline: [26024907](#)]
6. Salmon DA, Haber M, Gangarosa EJ, Phillips L, Smith NJ, Chen RT. Health consequences of religious and philosophical exemptions from immunization laws: individual and societal risk of measles. *JAMA* 1999 Jul 7;282(1):47-53. [Medline: [10404911](#)]
7. Hinman AR, Orenstein WA, Williamson DE, Darrington D. Childhood immunization: laws that work. *J Law Med Ethics* 2002;30(3 Suppl):122-127. [Medline: [12508514](#)]
8. Williams SE. What are the factors that contribute to parental vaccine-hesitancy and what can we do about it? *Hum Vaccin Immunother* 2014;10(9):2584-2596. [doi: [10.4161/hv.28596](#)] [Medline: [25483505](#)]
9. Etkind P, Lett SM, Macdonald PD, Silva E, Peppe J. Pertussis outbreaks in groups claiming religious exemptions to vaccinations. *Am J Dis Child* 1992 Feb;146(2):173-176. [Medline: [1733146](#)]
10. Fine P, Eames K, Heymann DL. “Herd immunity”: a rough guide. *Clin Infect Dis* 2011 Apr 1;52(7):911-916 [FREE Full text] [doi: [10.1093/cid/cir007](#)] [Medline: [21427399](#)]
11. Omer SB, Salmon DA, Orenstein WA, deHart MP, Halsey N. Vaccine refusal, mandatory immunization, and the risks of vaccine-preventable diseases. *N Engl J Med* 2009 May 7;360(19):1981-1988. [doi: [10.1056/NEJMsa0806477](#)] [Medline: [19420367](#)]

12. Dunn AG, Leask J, Zhou X, Mandl KD, Coiera E. Associations between exposure to and expression of negative opinions about human papillomavirus vaccines on social media: an observational study. *J Med Internet Res* 2015;17(6):e144 [FREE Full text] [doi: [10.2196/jmir.4343](https://doi.org/10.2196/jmir.4343)] [Medline: [26063290](https://pubmed.ncbi.nlm.nih.gov/26063290/)]
13. Radzikowski J, Stefanidis A, Jacobsen KH, Croitoru A, Crooks A, Delamater PL. The measles vaccination narrative in Twitter: a quantitative analysis. *JMIR Public Health Surveill* 2016 Jan 04;2(1):e1. [doi: [10.2196/publichealth.5059](https://doi.org/10.2196/publichealth.5059)] [Medline: [27227144](https://pubmed.ncbi.nlm.nih.gov/27227144/)]
14. Shaffer VA, Owens J, Zikmund-Fisher BJ. The effect of patient narratives on information search in a Web-based breast cancer decision aid: an eye-tracking study. *J Med Internet Res* 2013;15(12):e273 [FREE Full text] [doi: [10.2196/jmir.2784](https://doi.org/10.2196/jmir.2784)] [Medline: [24345424](https://pubmed.ncbi.nlm.nih.gov/24345424/)]
15. Kimmerle J, Bientzle M, Cress U. Personal experiences and emotionality in health-related knowledge exchange in Internet forums: a randomized controlled field experiment comparing responses to facts vs personal experiences. *J Med Internet Res* 2014 Dec;16(12):e277 [FREE Full text] [doi: [10.2196/jmir.3766](https://doi.org/10.2196/jmir.3766)] [Medline: [25486677](https://pubmed.ncbi.nlm.nih.gov/25486677/)]
16. Mazzocut M, Truccolo I, Antonini M, Rinaldi F, Omero P, Ferrarin E, et al. Web conversations about complementary and alternative medicines and cancer: content and sentiment analysis. *J Med Internet Res* 2016;18(6):e120 [FREE Full text] [doi: [10.2196/jmir.5521](https://doi.org/10.2196/jmir.5521)] [Medline: [27311444](https://pubmed.ncbi.nlm.nih.gov/27311444/)]
17. Ahluwalia A, Huang A, Bandari R, Roychowdhury V. An Automated Multiscale Map of Conversations: Mothers and Matters. *Social Informatics 4th International Conference*. Berlin: Springer Verlag; 2012:15-28. Presented at: SocInfo 2012; December 5-7, 2012; Lausanne, Switzerland. [doi: [10.1007/978-3-642-35386-4\\_2](https://doi.org/10.1007/978-3-642-35386-4_2)]
18. Bernhardt JM, Felter EM. Online pediatric information seeking among mothers of young children: results from a qualitative study using focus groups. *J Med Internet Res* 2004 Mar 01;6(1):e7 [FREE Full text] [doi: [10.2196/jmir.6.1.e7](https://doi.org/10.2196/jmir.6.1.e7)] [Medline: [15111273](https://pubmed.ncbi.nlm.nih.gov/15111273/)]
19. Blei DM, Ng AY, Jordan MI. Latent dirichlet allocation. *Journal of Machine Learning Research* 2003 Jan;3:993-1022 [FREE Full text]
20. Zhou Z. *Information Dynamics in Social Interactions: Hidden Structure Discovery and Empirical Case Studies*. Los Angeles: Electrical Engineering Dept., UCLA; 2013.
21. Greimas A. *Structural Semantics: An Attempt at a Method*. Lincoln: University of Nebraska Press; 1983.
22. Kitta A. *Vaccinations and Public Concern in History: Legend, Rumor, and Risk Perception*. New York: Routledge; 2012.
23. Tangherlini TR. "Who ya gonna call?": ministers and the mediation of ghostly threat in Danish legend tradition.. *Western Folklore* 1998;57(2/3):153-178. [doi: [10.2307/1500218](https://doi.org/10.2307/1500218)]
24. Labov W, Waletzky J. Narrative analysis: oral versions of personal experience. *J Narrative and Life Hist* 1997;7(1-4):3-38. [doi: [10.1075/jnlh.7.02nar](https://doi.org/10.1075/jnlh.7.02nar)]
25. Fader A, Soderland S, Etzioni O. Identifying relations for open information extraction. In: *Proceedings of the Conference on Empirical Methods in Natural Language Processing*. Stroudsburg: Association for Computational Linguistics; 2011 Presented at: Conference on Empirical Methods in Natural Language Processing; July 2011; Edinburgh.
26. Bigi B. Using Kullback-Leibler distance for text categorization. In: *Advances in Information Retrieval - 25th European Conference on IR Research*. Berlin: Springer; 2003 Presented at: European Conference on Information Retrieval; April 14, 2003; Pisa p. 305-319. [doi: [10.1007/3-540-36618-0\\_22](https://doi.org/10.1007/3-540-36618-0_22)]
27. Steven B. NLTK: the natural language toolkit. In: *COLING-ACL '06*. Stroudsburg: Association for Computational Linguistics; 2006 Presented at: COLING/ACL on Interactive presentation sessions; July 2006; Sydney p. 69-72. [doi: [10.3115/1225403.1225421](https://doi.org/10.3115/1225403.1225421)]
28. Levi-Strauss C. The structural study of myth. *J Am Folklore* 1955 Oct;68(270):428-444. [doi: [10.2307/536768](https://doi.org/10.2307/536768)]
29. Grün B, Hornik K. topicmodels: an R package for fitting topic models. *J Stat Soft* 2011;40(13):1-30. [doi: [10.18637/jss.v040.i13](https://doi.org/10.18637/jss.v040.i13)]
30. Ernst KC, Jacobs ET. Implications of philosophical and personal belief exemptions on re-emergence of vaccine-preventable disease: the role of spatial clustering in under-vaccination. *Hum Vaccin Immunother* 2012 Jun;8(6):838-841. [doi: [10.4161/hv.19743](https://doi.org/10.4161/hv.19743)] [Medline: [22508414](https://pubmed.ncbi.nlm.nih.gov/22508414/)]
31. Sun N, Rau PP, Ma L. Understanding lurkers in online communities: a literature review. *Comput Human Behav* 2014 Sep;38:110-117. [doi: [10.1016/j.chb.2014.05.022](https://doi.org/10.1016/j.chb.2014.05.022)]
32. Rodriguez N, Bollen J, Ahn YY. Collective dynamics of belief evolution under cognitive coherence and social conformity. *PLoS One* 2016;1(11):e0165910. URL: <https://arxiv.org/abs/1509.01502> [doi: [10.1371/journal.pone.0165910](https://doi.org/10.1371/journal.pone.0165910)]
33. Horne Z, Powell D, Hummel JE, Holyoak KJ. Countering antivaccination attitudes. *Proc Natl Acad Sci U S A* 2015 Aug 18;112(33):10321-10324 [FREE Full text] [doi: [10.1073/pnas.1504019112](https://doi.org/10.1073/pnas.1504019112)] [Medline: [26240325](https://pubmed.ncbi.nlm.nih.gov/26240325/)]
34. Hartley DM. Using social media and Internet data for public health surveillance: the importance of talking. *Milbank Q* 2014 Mar;92(1):34-39 [FREE Full text] [doi: [10.1111/1468-0009.12039](https://doi.org/10.1111/1468-0009.12039)] [Medline: [24597554](https://pubmed.ncbi.nlm.nih.gov/24597554/)]
35. Velasco E, Agheneza T, Denecke K, Kirchner G, Eckmanns T. Social media and Internet-based data in global systems for public health surveillance: a systematic review. *Milbank Q* 2014 Mar;92(1):7-33 [FREE Full text] [doi: [10.1111/1468-0009.12038](https://doi.org/10.1111/1468-0009.12038)] [Medline: [24597553](https://pubmed.ncbi.nlm.nih.gov/24597553/)]
36. Kümpel A, Karnowski V, Keyling T. News sharing in social media: a review of current research on news sharing users, content, and networks. *Soc Media Soc* 2015;1(2):1-14. [doi: [10.1177/2056305115610141](https://doi.org/10.1177/2056305115610141)]

37. Mollema L, Harmsen IA, 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](https://doi.org/10.2196/jmir.3863)] [Medline: [26013683](https://pubmed.ncbi.nlm.nih.gov/26013683/)]
38. Ozan-Rafferty ME, Johnson JA, Shah GH, Kursun A. In the words of the medical tourist: an analysis of Internet narratives by health travelers to Turkey. *J Med Internet Res* 2014;16(2):e43 [FREE Full text] [doi: [10.2196/jmir.2694](https://doi.org/10.2196/jmir.2694)] [Medline: [24513565](https://pubmed.ncbi.nlm.nih.gov/24513565/)]
39. Mazzocut M, Truccolo I, Antonini M, Rinaldi F, Omero P, Ferrarin E, et al. Web conversations about complementary and alternative medicines and cancer: content and sentiment analysis. *J Med Internet Res* 2016;18(6):e120 [FREE Full text] [doi: [10.2196/jmir.5521](https://doi.org/10.2196/jmir.5521)] [Medline: [27311444](https://pubmed.ncbi.nlm.nih.gov/27311444/)]
40. Chen AT, Zhu S, Conway M. What online communities can tell us about electronic cigarettes and hookah use: a study using text mining and visualization techniques. *J Med Internet Res* 2015;17(9):e220 [FREE Full text] [doi: [10.2196/jmir.4517](https://doi.org/10.2196/jmir.4517)] [Medline: [26420469](https://pubmed.ncbi.nlm.nih.gov/26420469/)]
41. Priest C, Knopf A, Groves D, Carpenter JS, Furrey C, Krishnan A, et al. Finding the patient's voice using big data: analysis of users' health-related concerns in the ChaCha Question-and-Answer Service (2009-2012). *J Med Internet Res* 2016;18(3):e44 [FREE Full text] [doi: [10.2196/jmir.5033](https://doi.org/10.2196/jmir.5033)] [Medline: [26960745](https://pubmed.ncbi.nlm.nih.gov/26960745/)]
42. Nicholson MS, Leask J. Lessons from an online debate about measles-mumps-rubella (MMR) immunization. *Vaccine* 2012 May 28;30(25):3806-3812. [doi: [10.1016/j.vaccine.2011.10.072](https://doi.org/10.1016/j.vaccine.2011.10.072)] [Medline: [22063388](https://pubmed.ncbi.nlm.nih.gov/22063388/)]
43. Kandadai V, Yang H, Jiang L, Yang CC, Fleisher L, Winston FK. Measuring health information dissemination and identifying target interest communities on Twitter: methods development and case study of the @SafetyMD Network. *JMIR Res Protoc* 2016;5(2):e50 [FREE Full text] [doi: [10.2196/resprot.4203](https://doi.org/10.2196/resprot.4203)] [Medline: [27151100](https://pubmed.ncbi.nlm.nih.gov/27151100/)]
44. Ayers JW, Westmaas JL, Leas EC, Benton A, Chen Y, Dredze M, et al. Leveraging big data to improve health awareness campaigns: a novel evaluation of the Great American Smokeout. *JMIR Public Health Surveill* 2016;2(1):e16 [FREE Full text] [doi: [10.2196/publichealth.5304](https://doi.org/10.2196/publichealth.5304)] [Medline: [27227151](https://pubmed.ncbi.nlm.nih.gov/27227151/)]
45. Surian D, Nguyen DQ, Kennedy G, Johnson M, Coiera E, Dunn AG. Characterizing Twitter discussions about HPV vaccines using topic modeling and community detection. *J Med Internet Res* 2016;18(8):e232 [FREE Full text] [doi: [10.2196/jmir.6045](https://doi.org/10.2196/jmir.6045)] [Medline: [27573910](https://pubmed.ncbi.nlm.nih.gov/27573910/)]
46. Bundorf MK, Wagner TH, Singer SJ, Baker LC. Who searches the Internet for health information? *Health Serv Res* 2006 Jun;41(3 Pt 1):819-836 [FREE Full text] [doi: [10.1111/j.1475-6773.2006.00510.x](https://doi.org/10.1111/j.1475-6773.2006.00510.x)] [Medline: [16704514](https://pubmed.ncbi.nlm.nih.gov/16704514/)]
47. Oraby T, Thampi V, Bauch CT. The influence of social norms on the dynamics of vaccinating behaviour for paediatric infectious diseases. *Proc Biol Sci* 2014 Apr 7;281(1780):20133172 [FREE Full text] [doi: [10.1098/rspb.2013.3172](https://doi.org/10.1098/rspb.2013.3172)] [Medline: [24523276](https://pubmed.ncbi.nlm.nih.gov/24523276/)]
48. Tang J, Gao H, Hu X, Liu H. Exploiting homophily effect for trust prediction. In: WSDM '13. New York: ACM; 2013 Feb Presented at: Sixth ACM international conference on Web search and data mining; February 2013; Rome p. 53-62. [doi: [10.1145/2433396.2433405](https://doi.org/10.1145/2433396.2433405)]
49. 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](https://doi.org/10.1177/1524839911405850)] [Medline: [21558472](https://pubmed.ncbi.nlm.nih.gov/21558472/)]

*Edited by G Eysenbach; submitted 30.08.16; peer-reviewed by A Dunn, L Grant; comments to author 20.09.16; revised version received 03.10.16; accepted 22.10.16; published 22.11.16.*

*Please cite as:*

Tangherlini TR, Roychowdhury V, Glenn B, Crespi CM, Bandari R, Wadia A, Falahi M, Ebrahimzadeh E, Bastani R  
"Mommy Blogs" and the Vaccination Exemption Narrative: Results From A Machine-Learning Approach for Story Aggregation on Parenting Social Media Sites  
*JMIR Public Health Surveill* 2016;2(2):e166  
URL: <http://publichealth.jmir.org/2016/2/e166/>  
doi:[10.2196/publichealth.6586](https://doi.org/10.2196/publichealth.6586)  
PMID:[27876690](https://pubmed.ncbi.nlm.nih.gov/27876690/)

©Timothy R Tangherlini, Vwani Roychowdhury, Beth Glenn, Catherine M Crespi, Roja Bandari, Akshay Wadia, Misagh Falahi, Ehsan Ebrahimzadeh, Roshan Bastani. Originally published in *JMIR Public Health and Surveillance* (<http://publichealth.jmir.org>), 22.11.2016. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in *JMIR Public Health and Surveillance*, is properly cited. The complete bibliographic information, a link to the original publication on <http://publichealth.jmir.org>, as well as this copyright and license information must be included.

Original Paper

# A Comparison of Web and Telephone Responses From a National HIV and AIDS Survey

Marcella K Jones<sup>1</sup>, MPH; Liviana Calzavara<sup>2,3</sup>, PhD; Dan Allman<sup>2,3</sup>, PhD; Catherine A Worthington<sup>2,4</sup>, PhD; Mark Tyndall<sup>2,5,6</sup>, MD, ScD, FRCPC; James Iveniuk<sup>2,3</sup>, PhD

<sup>1</sup>Faculty of Medicine, University of Toronto, Toronto, ON, Canada

<sup>2</sup>The CIHR Social Research Centre in HIV Prevention, Toronto, ON, Canada

<sup>3</sup>Dalla Lana School of Public Health, University of Toronto, Toronto, ON, Canada

<sup>4</sup>School of Public Health and Social Policy, University of Victoria, Victoria, BC, Canada

<sup>5</sup>BC Centre for Disease Control, Vancouver, BC, Canada

<sup>6</sup>Faculty of Medicine, University of British Columbia, Vancouver, BC, Canada

**Corresponding Author:**

Marcella K Jones, MPH

Faculty of Medicine

University of Toronto

Medical Sciences Building, Rm 2124, 1 King's College Circle

Toronto, ON, M5S 1A8

Canada

Phone: 1 416 946 8720

Fax: 1 416 946 3991

Email: [marcella.jones@mail.utoronto.ca](mailto:marcella.jones@mail.utoronto.ca)

## Abstract

**Background:** Response differences to survey questions are known to exist for different modes of questionnaire completion. Previous research has shown that response differences by mode are larger for sensitive and complicated questions. However, it is unknown what effect completion mode may have on HIV and AIDS survey research, which addresses particularly sensitive and stigmatized health issues.

**Objectives:** We seek to compare responses between self-selected Web and telephone respondents in terms of social desirability and item nonresponse in a national HIV and AIDS survey.

**Methods:** A survey of 2085 people in Canada aged 18 years and older was conducted to explore public knowledge, attitudes, and behaviors around HIV and AIDS in May 2011. Participants were recruited using random-digit dialing and could select to be interviewed on the telephone or self-complete through the Internet. For this paper, 15 questions considered to be either sensitive, stigma-related, or less-sensitive in nature were assessed to estimate associations between responses and mode of completion. Multivariate regression analyses were conducted for questions with significant ( $P \leq .05$ ) bivariate differences in responses to adjust for sociodemographic factors. As survey mode was not randomly assigned, we created a propensity score variable and included it in our multivariate models to control for mode selection bias.

**Results:** A total of 81% of participants completed the questionnaire through the Internet, and 19% completed by telephone. Telephone respondents were older, reported less education, had lower incomes, and were more likely from the province of Quebec. Overall, 2 of 13 questions assessed for social desirability and 3 of 15 questions assessed for item nonresponse were significantly associated with choice of mode in the multivariate analysis. For social desirability, Web respondents were more likely than telephone respondents to report more than 1 sexual partner in the past year (fully adjusted odds ratio (OR)=3.65, 95% CI 1.80-7.42) and more likely to have donated to charity in the past year (OR=1.63, 95% CI 1.15-2.29). For item nonresponse, Web respondents were more likely than telephone respondents to have a missing or "don't know" response when asked about: the disease they were most concerned about (OR=3.02, 95% CI 1.67-5.47); if they had ever been tested for HIV (OR=8.04, 95% CI 2.46-26.31); and when rating their level of comfort with shopping at grocery store if the owner was known to have HIV or AIDS (OR=3.11, 95% CI 1.47-6.63).

**Conclusion:** Sociodemographic differences existed between Web and telephone respondents, but for 23 of 28 questions considered in our analysis, there were no significant differences in responses by mode. For surveys with very sensitive health content, such



as HIV and AIDS, Web administration may be subject to less social desirability bias but may also have greater item nonresponse for certain questions.

(*JMIR Public Health Surveill* 2016;2(2):e37) doi:[10.2196/publichealth.5184](https://doi.org/10.2196/publichealth.5184)

## KEYWORDS

survey methods; cross-sectional survey; community survey; HIV; AIDS; social desirability; data collection methods; data quality; health attitude; health knowledge; attitudes; practice

## Introduction

Choosing a mode of data collection is one of the most important decisions that researchers make when designing a research survey. The choice of mode may be shaped by feasibility, cost, response rates, and potential for achieving a more representative sample. Currently, two of the most frequently used administration modes are telephone and Web [1,2]; the former is often interviewer-administered and the latter is usually self-administered. Often, telephone questionnaires have higher response rates and can also allow interviewers to clarify questions and motivate participants. However, telephone questionnaires can be relatively resource intensive and may have difficulty using visual aids [3,4]. By contrast, Web questionnaires are often less expensive and easier to implement but may be affected by lower response rates, item nonresponse, greater potential for fraud, and Internet accessibility issues that affect representativeness [3,5-7]. Given the compromises between modes, another option is a mixed-mode survey, which is administered using two or more data collection modes to allow participants an element of convenience and choice and to also compensate for the limitations of a single mode [1,2,8-10]. In each case, the choice of mode, or modes, may influence respondents' answers, and accordingly, the conclusions that researchers can draw about a study population.

One recurring concern in the literature is the relationship between survey mode and the accuracy of the information provided by the respondent because survey mode may influence respondents' willingness to answer certain questions at all, truthfully, or with socially desirable but inaccurate answers. Social desirability bias may be heightened in interviewer-administered telephone and face-to-face surveys but reduced in self-administered Web surveys [11,12]. Furthermore, although Web surveys typically have lower response rates than other modes of survey research, they can have lower item nonresponse rates as well, possibly because of the absence of interviewers whose characteristics can influence the kinds of attitudes and behaviors that people report [12-14]. In the context of mixed-mode surveys, these challenges can lead to inaccurate conclusions if researchers aggregate data that were collected through different modes [10], or if they compare results across multiple surveys that were collected by different modes [12]. In addition, if respondents are given the choice of mode, respondents may select a mode where the provision of socially desirable answers is more prevalent, or choose to not respond to certain questions, which may also prevent researchers from making accurate comparisons across subgroups [15,16].

The emergence of Web surveys has provided researchers with technology that creates new opportunities and challenges for

addressing response bias. Accordingly, there has been much recent research on Web surveys and how mode effects may shape responses to survey questions [5,9,10,17-23]. Although most research shows that social desirability bias is lower in Web surveys compared with other modes [4,12,24], other work has shown no difference [25,26]. This suggests that Web surveys may reduce social desirability bias only under certain circumstances and potentially only with certain types of questions. Alternatively, as Internet use is increasingly ubiquitous in society, the literature may be starting to reflect the fact that Web surveys may no longer evoke differential responses from survey participants.

In particular, how willing respondents are to report personal, private, or sensitive matters in a survey setting can affect data quality [9,17,18,22,27,28]. In a recent meta-analysis of 10 experimental studies that looked at Web administration versus interviewer-administered modes for collecting potentially embarrassing information, the authors concluded that self-administration via the Web improved reporting accuracy for socially undesirable responses [18]. Similarly, recent findings from the British National Survey of Sexual Attitudes and Lifestyles compared computer-assisted personal interviews with self-interview modes and found that reporting of sensitive information was overall higher in the Web/self-administered survey mode [21]. This could make Web surveys the preferred mode for asking highly sensitive personal questions.

This body of research is encouraging for research fields that investigate potentially sensitive topics. HIV and AIDS-related research often addresses particularly sensitive issues, and avoiding sensitive questions is not possible in most HIV- or AIDS-related surveys. Given the sensitive content, it is likely that studies of HIV and AIDS face particularly strong challenges arising from reporting biases, including social desirability and item nonresponse biases. To our knowledge, no national population-based experimental studies have looked at response differences between telephone and Web questionnaires for extremely sensitive health content, such as HIV and AIDS. Therefore, in this area of health research, observational studies that do not randomize participants to a completion mode can provide a helpful foundation for which to further explore the issue of response bias by mode. Indeed, it may be advantageous for health researchers considering mixed-mode survey designs to learn whether their specific topic is prone to response bias by mode, even when that data are nonexperimental in nature, such as 2 recent observational studies that explored response biases for specific health issues [26,27]. Understanding the extent to which responses may differ and how they differ depending on mode is essential given the use of mixed-mode



designs and for the comparability of studies on sensitive health topics that use different data collection modes.

A national HIV and AIDS survey was conducted among Canadians in 2011 to determine public knowledge, attitudes, and behaviors about HIV and AIDS. The survey used a mixed-mode data collection method, whereby participants selected whether they would prefer to complete the questionnaire by telephone or the Internet, which allows us to compare responses between (persons who chose) telephone and Web completion modes. We seek to determine whether the mode of questionnaire completion influences responses in terms of social desirability and missing data. To explore whether any observed associations vary by question type, we select different types of questions including potentially sensitive, stigma-related, and less-sensitive questions.

## Methods

### Questionnaire Development

The bilingual questionnaire was developed based on both literature reviews and the expertise of researchers and other professionals at the Canadian Institutes of Health Research (CIHR) Social Research Centre in HIV Prevention and the Canadian Foundation for AIDS Research. Questions were developed to resemble previous large-scale national HIV surveys for comparative purposes [29]. The final questionnaire contained sociodemographic items and questions about HIV and AIDS knowledge, attitudes, and behaviors (for more details, see [30]). Due to the study's sensitive nature, the survey was pretested among a sample (n=100) and monitored for issues (none were identified). Ethics approval was obtained from the University of Toronto Ethics Review Board.

### Survey Administration

The survey was conducted in English and French by The Strategic Counsel, a polling and market research firm, between May 5 and May 25, 2011 among individuals aged 16 years and older in all Canadian provinces and territories. Participants were selected using a 2-stage sampling design. In the first stage, participants were randomly sampled from the general population using a random-digit-dial that used both cellular and landline telephone numbers. Calls were managed by an interactive voice response system, with numbers retired from the system after 1 initial call and 3 unanswered callbacks. Once contacted, individuals entered basic sociodemographic information (age, gender, postal code) on their telephone keypads and were asked to participate in a survey at a later date. If the participant agreed, they were added to a panel of willing participants. For the second sampling stage, these panel participants comprised the sampling frame and were sampled directly (with stratification by region) and were contacted by a live interviewer who introduced the panel's incentive scheme and invited participants to complete the questionnaire over the Internet or by telephone. The incentive scheme was used to exclude professional

respondents and awarded participants with charity dollars (eg, option to donate to a charity of choice) and a ticket for a monthly raffle prize. The survey methodology is also described elsewhere [30-32]. The blended participation rate for the survey was 25%, with participation higher among those who completed the survey by telephone compared with those who did so through the Internet (31.1% and 18.4%, respectively). For this analysis, we exclude 16-17-year olds (n=54) because of their low Web-based completion (n=2).

### Measures

A total of 15 questions were selected from the 85 questions on the questionnaire to test differences in responses and were categorized by question type (Table 1). This subset of questions was used to minimize the effect of multiple comparisons testing. We chose 5 sensitive and 5 stigma-related questions that we hypothesized might be affected by response biases. We also chose 5 relatively benign or less-sensitive questions as controls. We use the terms "sensitive," "stigma-related," and "less-sensitive" as descriptors for the question types to organize our approach and interpretation; however, study participants were not aware of these categories and will have uniquely interpreted the sensitivity of each question.

As outlined in Table 1, to study differences in social desirability, we tested 13 questions that were suitable for predicting a socially desirable response. Two of the questions were excluded from this analysis because we did not believe there was a "socially correct" response, and if a mode difference were to be found, it would be difficult to interpret the directionality of the difference with respect to social desirability bias. Most questions had binary (yes or no) or Likert scale responses, except for question 1 (HIV or AIDS testing), where answers were recoded as voluntary versus nonvoluntary/nontested (using the same definition as in [32]), and for question 3 (number of sexual partners), where answers were recoded as 1 versus greater than 1. To study differences in missing data (eg, item nonresponse), all 15 questions were used. For analytical purposes, responses were dichotomized: missing or not missing. Both the "do not know/not sure" and "prefer not to answer" responses were grouped with missing responses because they were coded as missing during data entry. Complete details about the questions, including the response keys, are available in the Web-based Multimedia Appendix 1.

Our independent variable of interest, mode of completion, is a binary variable generated from each respondent's mode of questionnaire completion—web or telephone. Additional independent variables were self-reported sociodemographic items; these included age, gender, highest level of education attained, household income, sexual minority status, visible minority status, and Canadian region, all of which were treated as categorical variables in the analysis. Minority status was defined by the respondent indicating that they belonged to a "visible minority" group or a "sexual minority" group.

**Table 1.** An overview of the questions selected for analysis by question type and study objective (complete details about the questions are available in Web-based [Multimedia Appendix 1](#)).

Question type <sup>a</sup>	#	Question	Objective #1	Objective #2
			Social desirability	Item nonresponse
<b>Sensitive</b>				
	1	Have you ever been tested for HIV or AIDS for any of the following reasons?	✓	✓
	2	Have you ever had sex in your lifetime?	✓	✓
	3	How many different partners have you had sexual intercourse with in the last 12 months? <sup>b</sup>	✓	✓
	4	Were any of these casual partners? In other words, were they someone that you are not in a regular or long-term relationship? <sup>b</sup>	✓	✓
	5	What is your annual household income from all sources before taxes? <sup>c</sup>		✓
<b>Stigma-related</b>				
	6	I could not become friends with someone who has HIV or AIDS (select level of agreement with statement)	✓	✓
	7	I feel afraid of people living with HIV or AIDS (select level of agreement with statement)	✓	✓
	8	People living with HIV or AIDS have the right to be sexually active (select level of agreement with statement)	✓	✓
	9	How comfortable or uncomfortable would you be with a close friend or family member dating someone with HIV or AIDS? (select level of comfort)	✓	✓
	10	How comfortable or uncomfortable would you be with shopping at a small neighborhood grocery store, if you found out that the owner had HIV or AIDS? (select level of comfort)	✓	✓
<b>Less-sensitive</b>				
	11	Thinking about illnesses or diseases, what is the one illness or disease that concerns you the most? (open ended) <sup>c</sup>		✓
	12	In the past year, did you actively seek out or look for information about HIV or AIDS?	✓	✓
	13	Do you recall donating to any charitable or not-for-profit organization in the last year?	✓	✓
	14	To what extent do you believe it is government's responsibility to continue to fund HIV or AIDS research?	✓	✓
	15	How knowledgeable would you say that you are about HIV or AIDS?	✓	✓
	Total		13	15

<sup>a</sup>Question types are categories that are used to guide our analysis and may not be perceived this way by participants.

<sup>b</sup>Were only asked to those who were sexually active in the past 12 months.

<sup>c</sup>Were excluded from objective 1 because we did not believe there was a "socially correct" response.

## Statistical Analyses

Analyses were performed with Stata IC v. 12 using its survey data analysis program. All reported results are weighted to represent the Canadian population in 2011 in terms of age, gender, and province or territory of residence. Standard errors were estimated using linearized or robust variance estimators, and 95% CIs are presented where appropriate. Descriptive statistics and bivariate associations with mode of completion were generated for all sociodemographic variables. Bivariate associations between mode of completion and the selected questions were generated using Pearson's chi-square or Wald tests as appropriate.

Any significant bivariate associations observed were either deemed attributable to sample differences in the study population between the telephone and Web groups or to the completion mode itself. To address the first possibility, sample differences between mode groups, a multivariate analysis was conducted. Multivariate regression analyses were conducted for only those questions with significant differences ( $P \leq .05$ ) in responses or missing responses between telephone and Web administration modes. Separate regressions were run for each question: logistic regression was used for binary responses and linear regression for continuous responses. First, sociodemographic variables with significant bivariate associations ( $P \leq .05$ ) with mode of administration were included

in the regression models as control variables. Second, to minimize the effect of mode selection bias on confounding our results, we use propensity score methodology, which was developed to approximate the analysis of observational (nonrandomized) data to that of randomized treatment assignment [33]. The propensity score balances systematic differences between the telephone and Web response groups so that observed sociodemographic covariates are similar between the 2 groups [33]. Therefore, the inclusion of the propensity score as a covariate in our multivariate analysis helps reduce bias that may be present as a result of respondents' self-selection into telephone or Web response modes. We generated a propensity score using a logistic regression model in which mode of completion was regressed on all the observed sociodemographic characteristics in our study (age, gender, education, income, sexual minority, ethnic minority, and province of residence). An individual's estimated propensity score is therefore the predicted probability of that individual choosing to complete their survey on the Internet. After the propensity score variable was generated, it was added as a covariate in our multivariate models, by simply including it as an independent variable in the model statement.

Respondents with any missing observations for covariates were excluded from the regression analysis, with the exception of missing household income values where an additional response category was generated to maintain sample size. All logistic regression models satisfied the Hosmer–Lemeshow test for goodness of fit ( $F$ -adjusted mean residual  $>.05$ ) [34].

## Results

### Characteristics of the Study Population

In total, 2085 Canadians aged 18 years and older completed the questionnaire. Overall, the unweighted sample closely reflected the actual distribution of the 2011 Canadian population in terms of key demographic variables such as age, gender, and province or territory (Table 2). The study sample is more highly educated, however, than the general Canadian population. A total of 1690 participants (81.0%) completed the questionnaire through the Internet, and 395 (19.0%) completed by telephone. Table 3 presents the sociodemographic characteristics of the study sample by mode of questionnaire completion. Mode was significantly associated with most of the sociodemographic characteristics we considered, except for visible minority status. Compared with Web respondents, telephone respondents tended to be older ( $P<.001$ ), were more likely to be female ( $P=.05$ ), reported less education ( $P<.001$ ), had lower household incomes ( $P<.001$ ), and were more likely to live in the province of Quebec ( $P<.001$ ). Those who self-identified as being a member of a

sexual minority group were more likely to have chosen to complete the questionnaire through the Internet ( $P=.004$ ).

### Objective 1: Responses to Questions According to Mode of Completion

Responses for the 13 social desirability questions overall and by mode of questionnaire completion are presented in Table 4. Significant differences in responses were observed for 5 of the 13 questions. Among those who had been sexually active in the last 12 months, a significantly higher proportion of Web respondents (13.8%) than telephone respondents (4.3%) reported having more than 1 sexual partner in the last 12 months ( $P<.001$ ). Web respondents reported a higher level of comfort with shopping at a small neighborhood grocery store where the owner was known to have HIV or AIDS (3.20 vs 2.92 for telephone respondents,  $P<.001$ ). In response to whether survey participants agreed with the stigma-related statement “I feel afraid of people living with HIV and AIDS,” the overall mean level of agreement in the study population was low (2.53), which falls between “2—disagree” and “3—somewhat disagree.” Web respondents tended toward “disagree” (2.48), whereas telephone respondents tended toward “somewhat disagree” (2.74;  $P=.021$ ). Web respondents also reported more charitable giving in the past year (86.5%) compared with telephone respondents (77.4%;  $P<.001$ ) and greater self-reported knowledge of HIV or AIDS (4.56 vs 4.39 for telephone respondents,  $P=.046$ ). The remaining 8 questions showed no statistically significant differences in responses between the Web and telephone respondents.

### Objective 2: Missing Data According to Mode of Completion

Missing responses for the 15 questions overall and by mode of completion are summarized in Table 5. Overall, the frequency of missing data was low and ranged from 0.9% to 4.7% with the exception of 2 questions: annual household income (15.2% missing) and illness or disease that concerns you the most (9.8% missing). Mode did not affect refusing to report annual household income. Significant differences in missing responses were observed for 3 of the 12 questions. A missing or do not know response to whether the respondent had tested for HIV was significantly more likely in Web respondents (4.9%) than telephone respondents (0.7%;  $P<.001$ ). Web respondents were also more likely to have a missing or do not know response when asked to rate their comfort level with shopping at a grocery store owned by someone who has HIV or AIDS (5.2% vs 2.7% for telephone respondents,  $P=.041$ ) and when asked about what illness or disease concerns them the most (11.1% vs 4.3% for telephone respondents,  $P<.001$ ). No significant differences in missing responses between the modes were observed for the remaining 12 questions.

**Table 2.** Key variables from the 2011 general population in Canada, aged 18 years and older, compared with unweighted and weighted survey samples (n=2085).

Variable	Categories	Canadian Population <sup>a</sup> %	Unweighted survey sample %	Weighted survey sample %
Age				
	18-24	11.5	10.1	12.1
	25-39	24.4	26.7	25.0
	40-49	18.8	16.9	16.8
	50-59	18.8	19.5	19.2
	60+	26.3	26.9	26.9
Gender				
	Male	48.5	49.3	48.2
	Female	51.5	50.7	51.8
Education <sup>b</sup>				
	Male	48.5	49.3	48.2
	Female	51.5	50.7	51.8
	University	16.5	28.0	28.0
	Grad or postgrad	9.4	17.3	17.2
Region				
	British Columbia	13.4	13.0	12.3
	ALberta	10.6	10.1	10.2
	Saskatchewan	3.0	3.7	3.0
	Manitoba	3.5	3.9	3.5
	Ontario	38.3	36.8	38.3
	Quebec	23.9	23.5	24.0
	New Brunswick	2.3	2.0	2.3
	Nova Scotia	2.8	4.1	3.2
	Prince Edward Island	0.4	0.4	0.3
	Newfoundland and Labrador	1.6	1.3	1.6
	Yukon	0.1	0.8	0.8
	Northwest	0.1	0.4	0.4
	Nunavut	0.1	0.1	0.1

<sup>a</sup>Based on 2011 census of Canada [35-37].

<sup>b</sup>Education shown for age 25 years and older (n=1860), as the Canadian Census does not report highest level of education for less than 25 years.

**Table 3.** Sociodemographics of study population by mode of questionnaire completion.

Variable	Categories	Overall <sup>a</sup> % (n=2085)	Web % (n=1690)	Telephone % (n=395)	P value <sup>b</sup>
Overall (n=2085)		100	81.0	19.0	
Age (n=2085)					
	18-24	12.1	13.1	7.9	<.001
	25-39	25.0	28.0	12.3	
	40-49	16.8	17.1	15.4	
	50-59	19.2	18.7	21.5	
	60+	26.9	23.1	42.9	
Gender (n=2085)					
	Male	48.2	49.3	43.8	.050
	Female	51.8	50.7	56.2	
Education (n=2069)					
	≤High school	20.8	14.3	48.3	<.001
	College	31.9	32.5	29.3	
	University	31.9	35.4	17.2	
	Grad or postgrad	15.5	17.8	5.3	
Income (n=2085)					
	<\$40,000	23.3	19.6	39.1	<.001
	\$40,000-\$80,000	28.4	28.1	29.6	
	>\$80,000	33.2	37.8	13.7	
	Prefer not to answer	15.2	14.6	17.7	
Sexual minority (n=2038)		4.8	5.5	2.0	.004
Visible minority (n=2038)		7.7	7.5	8.6	.462
Region (n=2085)					
	West	29.1	30.6	22.7	<.001
	Ontario	38.3	39.1	35.0	
	Quebec	24.0	21.6	34.1	
	East	7.4	7.3	8.1	
	Territories	1.3	1.5	0.2	

<sup>a</sup>Overall % based on “n” from cross tabulation.

<sup>b</sup>Pearson chi-square corrected for weighted data (design-based F).



**Table 4.** Objective 1 (social desirability): responses to questions by mode of questionnaire completion.

Question type	Question	Response or scale <sup>a</sup>	Overall <sup>b</sup> % or mean (standard error-SE)	Web % or mean (SE)	Telephone % or mean (SE)	<i>P</i> value <sup>c</sup>
Sensitive questions						
	Voluntary testing for HIV or AIDS (n=1997)	Yes	29.7	29.9	29.0	.739
	Sexual activity in lifetime (n=2054)	Yes	96.7	96.6	96.9	.735
	More than 1 sexual partner in last 12 months (n=1455) <sup>d</sup>	Yes	12.4	13.8	4.3	<.001
	Casual sex partners in last 12 months (n=1461) <sup>d</sup>	Yes	12.2	12.8	8.7	.089
	Could not become friends with someone who has HIV or AIDS (n=2055)	1-7	1.93 (0.04)	1.91 (0.04)	2.04 (0.09)	.169
	Afraid of people living with HIV or AIDS (n=2065)	1-7	2.53 (0.04)	2.48 (0.04)	2.74 (0.11)	.021
	People with HIV or AIDS have the right to be sexually active (n=2005)	1-7	4.39 (0.05)	4.35 (0.05)	4.57 (0.11)	.084
	Comfort with a close friend or family member dating someone with HIV or AIDS (n=2008)	1-4	2.37 (0.02)	2.37 (0.03)	2.38 (0.05)	.897
	Comfort with shopping at small grocery store owned by someone who has HIV or AIDS (n=1987)	1-4	3.15 (0.02)	3.20 (0.02)	2.92 (0.06)	<.001
Less-sensitive questions						
	Actively seek information about HIV or AIDS (n=2061)	Yes	8.3	8.4	8.2	.919
	Charitable giving in the past year (n=2049)	Yes	84.7	86.5	77.4	<.001
	Government responsibility to fund HIV or AIDS research (n=2053)	1-7	5.54 (0.04)	5.51 (0.04)	5.65 (0.09)	.136
	Perceived HIV knowledge (n=2080)	1-7	4.52 (0.03)	4.56 (0.03)	4.39 (0.08)	.046

<sup>a</sup>1-7 Likert scale; 1=completely disagree, 4=neither agree nor disagree, and 7=completely agree. 1-4 Likert scale; 1=very uncomfortable, and 4=very comfortable.

<sup>b</sup>Overall % or mean based on "n" from cross tabulation.

<sup>c</sup>Pearson chi-square (categorical variables) or Wald test (continuous variables) corrected for weighted data (design-based F).

<sup>d</sup>These questions were only asked to those who were sexually active in the past 12 months.

**Table 5.** Objective 2 (item nonresponse): missing or do not know<sup>a</sup> responses for questions by mode of questionnaire completion (n=2085, unless otherwise indicated).

Question type	Question	Overall <sup>b</sup> %	Web %	Telephone %	P value <sup>c</sup>
Sensitive questions					
	Tested for HIV or AIDS	4.1	4.9	0.7	<.001
	Sexual activity in lifetime	3.4	3.6	2.3	.212
	Number of sexual partners in last 12 months <sup>d</sup>	1.3	1.4	0.8	.476
	Casual sex partners in last 12 months <sup>d</sup>	0.9	1.1	0.0	.127
	Annual household income	15.2	14.6	17.7	.129
Stigma-related questions					
	Could not become friends with someone who has HIV or AIDS	1.4	1.7	0.5	.082
	Afraid of people living with HIV or AIDS	1.0	0.8	1.8	.061
	People with HIV or AIDS have the right to be sexually active	3.8	3.7	4.3	.567
	Comfort with a close friend or family member dating someone with HIV or AIDS	3.4	3.9	3.4	.678
	Comfort with shopping at small grocery store owned by someone who has HIV or AIDS	4.7	5.2	2.7	.041
Less-sensitive questions					
	Illness or disease that concerns you the most	9.8	11.1	4.3	<.001
	Actively seek information about HIV or AIDS	1.1	1.3	0.3	.085
	Charitable giving in the past year	1.7	1.9	1.0	.249
	Government responsibility to fund HIV or AIDS research	1.5	1.6	1.0	.406
	Perceived HIV knowledge	0.2	0.2	0.3	.983

<sup>a</sup>All responses recategorized into binary variables: missing or do not know (yes) or other (no).

<sup>b</sup>Overall % based on "n" from cross tabulation.

<sup>c</sup>Pearson chi-square corrected for weighted data (design-based F).

<sup>d</sup>These questions were only asked to those who were sexually active in the past 12 months (n=1474).

### Findings from the Multivariate Analysis

The 8 questions that exhibited significant associations with mode in the bivariate analysis were entered into separate regression models to control for potential confounding by differences between the online and telephone samples. Adjusted associations with mode of completion and these questions are shown in Table 6. After adjusting for sociodemographic differences between the Web and telephone groups, differences in responses remain for 5 of the 8 questions. When the propensity score is added as a covariate to adjust for mode selection bias, the significant differences remain. While the propensity score does not resolve unmeasured differences between the two groups, the fully adjusted estimates are controlled for any confounding due to observed systematic differences that predict choice of response mode. Web respondents had 3.65 greater odds of reporting more than one

sexual partner in the last 12 months, compared to telephone respondents (sociodemographic and propensity score adjusted odds ratio (OR)=3.65, 95% CI 1.80-7.42). Those who completed online were also more likely than telephone respondents to report charitable giving in the past year (OR=1.63, 95% CI 1.15-2.29). In terms of item non-response, web respondents were significantly more likely to have a missing or "don't know" response to questions about HIV testing (OR=8.04, 95% CI 2.46-26.31), comfort with shopping at small grocery store owned by someone who has HIV or AIDS (OR=3.11, 95% CI 1.47-6.63), and most concerning illness or disease (OR=3.02, 95% CI 1.67-5.47). After multivariate adjustment, the previously observed significant differences in responses between the modes for the other three questions did not persist. This suggests that the bivariate associations for these questions can be attributed to sociodemographic differences between the samples.

**Table 6.** Adjusted associations<sup>a</sup> between mode of questionnaire completion (exposure) and 8 selected questions (outcomes; see [Multimedia Appendix 2](#) on the Internet for full regression results).

Question type and question	Mode	Sociodemographic adjusted <sup>b</sup>		Sociodemographic <sup>b</sup> and propensity score adjusted	
		Odds ratio or $\beta$ coefficient	95% CI	Odds ratio or $\beta$ coefficient	95% CI
<b>Objective 1: social desirability</b>					
Sensitive: More than 1 sexual partner in the last 12 months (n=1424)					
	Telephone	1.00	—	1.00	—
	Web	3.76 <sup>d</sup>	(1.86-7.59)	3.65 <sup>d</sup>	(1.80-7.42)
Stigma-related: afraid of people living with HIV or AIDS (n=2008) <sup>c</sup>					
	Telephone	0.00	—	0.00	—
	Web	0.000	(-0.23-0.23)	0.019	(-0.21-0.25)
Stigma-related: comfort with shopping at small grocery store owned by someone who has HIV or AIDS (n=1934) <sup>c</sup>					
	Telephone	0.00	—	0.00	—
	Web	0.087	(-0.34-0.21)	0.069	(-0.05-0.19)
Less-sensitive: Charitable giving in the past year (n=1996)					
	Telephone	1.00	—	1.00	—
	Web	1.61 <sup>d</sup>	(1.15-2.27)	1.63 <sup>d</sup>	(1.15-2.29)
Less-sensitive: perceived HIV knowledge (n=2022) <sup>c</sup>					
	Telephone	0.00	—	0.00	—
	Web	-0.049	(-0.22-0.12)	-0.057	(-0.23-0.11)
<b>Objective 2: Item Non-Response</b>					
Sensitive: tested for HIV or AIDS (n=2027)					
	Telephone	1.00	—	1.00	—
	Web	8.68 <sup>d</sup>	(2.63-28.67)	8.04 <sup>d</sup>	(2.46-26.31)
Stigma-related: comfort with shopping at small grocery store owned by someone who has HIV or AIDS (n=2027)					
	Telephone	0.00	—	0.00	—
	Web	2.99 <sup>d</sup>	(1.41-6.32)	3.11 <sup>d</sup>	(1.47-6.63)
Less-sensitive: illness or disease that concerns you the most (n=2027)					
	Telephone	1.00	—	1.00	—
	Web	3.01 <sup>d</sup>	(1.68-5.38)	3.02 <sup>d</sup>	(1.67-5.47)

<sup>a</sup>Separate logistic regression analyses were run for all questions except for continuous outcomes.

<sup>b</sup>Adjusted for age, gender, education, household income, sexual minority status, and region.

<sup>c</sup>Linear regression analysis.

<sup>d</sup>Indicates statistical significance at the 95% confidence level.

## Discussion

We sought to determine whether there are differences in social desirability and missing data between people who chose to complete an HIV- and AIDS-related questionnaire by telephone versus the Web. We anticipated that due to the sensitive and personal nature of some of the questions, we would see differences in responses between the telephone interviewer-administered questionnaire and the Web-based

self-completed questionnaire. Although we saw large sociodemographic differences between completion modes, for most of the questions (23 of 28) we studied, there were no significant differences in responses by mode. Overall, 2 of 13 questions assessed for social desirability, and 3 of 15 questions assessed for item nonresponse were significantly associated with choice of mode in the multivariate analysis. However, despite finding few significant response differences, those we found were large in magnitude; ranging from 1.6- to 8-fold difference.

Participants were given a choice to complete the questionnaire through the Internet or by telephone. As expected, and similar to other mixed-mode studies that used a non-random method for selecting participants [17,27,38], we see large differences in sociodemographic characteristics between the Web and telephone groups. The telephone group was older and had a larger proportion of women. Similar to other Canadian studies, we also found that Internet respondents reported higher incomes and more education [27,38]. Other studies have found younger age, higher incomes, greater education, and race to be strongly associated with the Internet and email access [6,39-41]. Although we do not know whether that telephone respondents in our study have access to the Internet, our results are consistent with what these findings suggest about Internet access. The exception is that we found no difference in the proportion of visible minorities between the two modes; this finding may be unique to the Canadian population as compared with the American population, owing to more targeted social policies in Canada to expand Internet access and reduce the digital divide for disadvantaged and racialized groups [42].

We thought that sensitive and stigma-related questions might be differentially affected by social desirability between completion modes, but a pattern in responses by question type was not apparent. We saw some evidence of social desirability for sensitive questions, with telephone respondents reporting lower numbers to questions about sexual partners they had in the past year and in answering whether any of these sexual partners were casual partners. This is in line with previous research that has found that Web questionnaires are better at eliciting truthful responses to sensitive questions than telephone questionnaires [9,20,22,27,28]. However, only the question about number of sexual partners showed a statistically significant difference, a difference that remained after multivariate adjustment. This suggests that the social interaction with the interviewer may have resulted in respondents stating a lower number of sexual partners—a more socially appropriate response.

Four other questions that we tested for social desirability showed significant differences in responses by mode, but not in the anticipated direction. If social desirability was at play, telephone respondents would be expected to provide more tolerant responses about their attitudes toward people living with HIV and AIDS because these attitudes are more socially appropriate. However, we found that telephone respondents gave less tolerant responses than Web respondents when asked if they felt afraid of people living with HIV and AIDS and when rating their comfort level with shopping at a grocery store owned by someone who is HIV seropositive. Yet, these differences did not remain in the multivariate analysis and are therefore attributed to sample differences between mode groups. Similarly, telephone respondents reported less knowledge about HIV and AIDS and less charitable giving than Web respondents, which are considered socially undesirable responses, yet only, the difference in charitable giving persisted in the multivariate analysis. This result suggests that when reporting their charitable donations to a live interviewer, telephone respondents were not affected by social desirability and did not feel socially obligated to appear more charitable. Although we adjusted for income

and education in our models, Web respondents may be different from telephone respondents in other important characteristics that could account for the counterintuitive result to this question, such as religious affiliation, awareness of need, altruism, and personal values, all of which are known motivators for charitable giving [43]. Another possible explanation is that some telephone respondents may fear that the live interviewer will ask them to donate to a charity at the end of the questionnaire if they appear charitable and therefore choose to answer “no” to avoid the perceived solicitation.

With respect to the missing data objective, we also saw no clear pattern of response differences between question types. Among the questions we considered, there was generally a greater frequency of item nonresponse to sensitive and stigmatizing questions and among the Web sample. One question from each category was found to have significant differences in item nonresponse between the telephone and Web groups, with a greater frequency of missing responses in the Web questionnaire. These differences also remained in the multivariate analysis, which suggests that they are attributable to the mode and not to sociodemographic differences between the samples. The greatest amount of item nonresponse was for annual household income (15.2%), with more nonresponse among telephone participants (17.7%) than Web participants (14.6%), although this difference was not statistically significant. Questions about income are well known to generate a large frequency of missing data in most surveys.

Our findings for item nonresponse are consistent with those of other experimental and observational studies that looked at missing data by mode and also found that Web questionnaires produced more missing data [9,27,44,45]. It is thought that telephone interviewers may inadvertently persuade respondents to answer, or that they may further explain or clarify questions. Some of these studies found that differences were particularly evident for complicated or difficult questions [9], although neither question that we saw differences for is considered difficult. Due to the nature of how the data were entered and coded, we are unable to distinguish between different types of missing data (eg, “do not know” responses vs not answered questions), and therefore, we cannot tell where “do not know” responses may be genuine answers. The concerning illness or disease question was an open-ended question, and this likely contributed to item nonresponse among the Web sample as typing was required, and no interviewer was present to prompt for a response. This is consistent with other studies that have also found increased missing data for open-ended questions on Web surveys [22,27]. In contrast, the HIV testing question provided response categories, but some respondents may not have known if they had ever been tested for HIV and legitimately selected the “do not know” response, which would have contributed to item nonresponse for this question.

Our study benefits from several strengths. We conducted a large national survey, and so, we anticipate that our results have fair generalizability to the Canadian population. Our sample was recruited from a respondent panel that was constructed using random-digit-dial, and this strengthens our ability to make causal inferences [46]. We also performed a multivariate regression analysis to control for sample differences between the telephone

and Web groups and included a propensity score that can help adjust for mode selection effects [10]. After accounting for sociodemographic disparities between the groups, it is more likely that any remaining differences in responses can be attributed to the mode of completion.

The results of our observational study should be considered along with its limitations. First, and most importantly, we did not randomize respondents to mode, and therefore, our estimates may be subject to selection bias. We attempted to control for the risk of selection bias by including a propensity score in our regression models. The inclusion of the propensity score neither changed the significance of our results from the sociodemographic-adjusted models nor did it greatly change their magnitude. This indicates that relatively little selection bias is present. Furthermore, the fact that our results persist in the face of the propensity score control would suggest that our results do describe the effect of survey mode on response patterns. However, there may be other unmeasured characteristics that we are unable to control for, such as differences in people's experiences and opinions between the groups, and if these characteristics also affect mode selection, this could be confounding our results. Although, if these characteristics are unrelated to mode selection, they would only produce nondifferential misclassification, which would indicate our estimates are conservative. Furthermore, the regression model that generated the propensity score was limited to variables contained within our questionnaire, and as such, there may be variables that help explain why one chose a given survey mode that are missing from the model specification. The propensity score is therefore not a perfect predictor of mode selection, and there is some degree of misspecification as the probability of choosing a mode and actually choosing it are distinct constructs. In addition, because the sample size of telephone respondents was much smaller than the sample of Web respondents, the propensity score may not be able to adequately balance sociodemographic variables across the response groups. Second, the participation rate from the random-digit-dial panel was low, and more people who opted to complete by telephone finished the questionnaire (31.1%) than those who opted to complete through the Internet (18.4%).

Although this participation rate limits the external validity of our results, particularly for the Web sample, it is nevertheless consistent with response rates from similar surveys in Canada that report response rates between 14% and 28% [31,47]. Finally, beyond social desirability and missing data, we did not consider other potential ways that responses could differ between modes, such as nondifferentiation bias (eg, answering the same response across a series of items), acquiescence bias (eg, agreeing across questions), and avidity bias (eg, disproportionate representation of those interested or invested in the survey topic of HIV and AIDS), due to study scope and design of the questionnaire.

In summary, there has been considerable concern in the survey methodology literature about the use of Web questionnaires to conduct research. Primary concerns include low response rates, low representativeness, sampling issues, and the comparability of Web-collected data to data collected by other modes [6,44,46,48]. Although these are important concerns that are not fully addressed by our study, we have shown that in a nonrandomized mixed-mode survey for a particularly sensitive health topic, the differences between telephone- and Web-collected responses were minimal in terms of social desirability bias and item nonresponse. This is promising given that Web surveys are significantly cheaper and faster to implement. On the basis of our analysis in this nonrandomized study, we cautiously suggest a comparative strength of Web surveys is that they may provide more truthful results, particularly for questions about sexual behaviors. Although we did see more item nonresponse among the Web group overall, it was mostly nonsignificant and may be addressed through improved Web questionnaire design, such as using prompts that alert respondents if they attempt to move forward without completing a question [2,22].

Our results, although limited by the observational study design, may be important to consider in light of the increasing use of mixed-mode surveys that combine telephone and Web completion modes to reduce costs and increase validity and may help inform future experimental survey methods studies and population-based research in the area of HIV and AIDS.

---

## Acknowledgments

This work was supported by a generous donation from Andy Pringle and in-kind support of the CIHR Social Research Centre in HIV Prevention (SRC).

The authors specially thank the study participants, The Strategic Counsel for conducting the survey, Canadian Foundation for AIDS Research (CANFAR) for their collaboration on the project, and SRC staff and members for their work on the national study (Laura Bisailon, Caroline Godbout, Jill Morse, Robin Montgomery, Melissa Walters and Samantha White). They also thank Rahim Moineddin for consulting on the statistical methods for this paper.

---

## Conflicts of Interest

None declared.

---

## Multimedia Appendix 1

[PDF File (Adobe PDF File), 48KB - [publichealth\\_v2i2e37\\_app1.pdf](#) ]



## Multimedia Appendix 2

[[PDF File \(Adobe PDF File\), 77KB - publikealth\\_v2i2e37\\_app2.pdf](#)]

### References

1. Blom AG, Bosnjak M, Cornilleau A, Cousteaux A, Das M, Douhou S, et al. A Comparison of Four Probability-Based Online and Mixed-Mode Panels in Europe. *Social Science Computer Review* 2016 Feb;34(1):8-25. [doi: [10.1177/0894439315574825](https://doi.org/10.1177/0894439315574825)]
2. Dillman D, Smyth J, Christian LM. *Internet, phone, mail, and mixed-mode surveys: the tailored design method*. New York: Wiley; 2014.
3. Jones TL, Baxter MA, Khanduja V. A quick guide to survey research. *Ann R Coll Surg Engl* 2013 Jan;95(1):5-7 [FREE Full text] [doi: [10.1308/003588413X13511609956372](https://doi.org/10.1308/003588413X13511609956372)] [Medline: [23317709](https://pubmed.ncbi.nlm.nih.gov/23317709/)]
4. Tourangeau R. Measurement Properties of Web Surveys. In: *Beyond traditional survey taking: adapting to a changing world.*: Statistics Canada; 2014 Presented at: Proc Stat Canada Symp; 2014; Gatineau, Québec, Canada.
5. Bauermeister J, Pingel E, Zimmerman M, Couper M, Carballo-Diéguez A, Strecher VJ. Data Quality in web-based HIV/AIDS research: Handling Invalid and Suspicious Data. *Field methods* 2012 Aug 1;24(3):272-291 [FREE Full text] [doi: [10.1177/1525822X12443097](https://doi.org/10.1177/1525822X12443097)] [Medline: [23180978](https://pubmed.ncbi.nlm.nih.gov/23180978/)]
6. Rivara FP, Koepsell TD, Wang J, Durbin D, Jaffe KM, Vavilala M, et al. Comparison of telephone with World Wide Web-based responses by parents and teens to a follow-up survey after injury. *Health Serv Res* 2011 Jun;46(3):964-981 [FREE Full text] [doi: [10.1111/j.1475-6773.2010.01236.x](https://doi.org/10.1111/j.1475-6773.2010.01236.x)] [Medline: [21275989](https://pubmed.ncbi.nlm.nih.gov/21275989/)]
7. Teitcher JE, Bockting WO, Bauermeister JA, Hofer CJ, Miner MH, Klitzman RL. Detecting, preventing, and responding to “fraudsters” in internet research: ethics and tradeoffs. *J Law Med Ethics* 2015;43(1):116-133 [FREE Full text] [doi: [10.1111/jlme.12200](https://doi.org/10.1111/jlme.12200)] [Medline: [25846043](https://pubmed.ncbi.nlm.nih.gov/25846043/)]
8. Bosnjak M, Das M, Lynn P. Methods for Probability-Based Online and Mixed-Mode Panels: Selected Recent Trends and Future Perspectives. *Social Science Computer Review* 2016 Feb;34(1):3-7. [doi: [10.1177/0894439315579246](https://doi.org/10.1177/0894439315579246)]
9. Greene J, Speizer H, Wiitala W. Telephone and web: mixed-mode challenge. *Health Serv Res* 2008 Feb;43(1 Pt 1):230-248 [FREE Full text] [doi: [10.1111/j.1475-6773.2007.00747.x](https://doi.org/10.1111/j.1475-6773.2007.00747.x)] [Medline: [18211527](https://pubmed.ncbi.nlm.nih.gov/18211527/)]
10. Hox JJ, De Leeuw ED, Zijlman EA. Measurement equivalence in mixed mode surveys. *Front Psychol* 2015 Feb;6:87 [FREE Full text] [doi: [10.3389/fpsyg.2015.00087](https://doi.org/10.3389/fpsyg.2015.00087)] [Medline: [25699002](https://pubmed.ncbi.nlm.nih.gov/25699002/)]
11. Persoskie A, Leyva B, Ferrer RA. Mode Effects in Assessing Cancer Worry and Risk Perceptions: Is Social Desirability Bias at Play? *Med Decis Making* 2014 Jul;34(5):583-589 [FREE Full text] [doi: [10.1177/0272989X14527173](https://doi.org/10.1177/0272989X14527173)] [Medline: [24718657](https://pubmed.ncbi.nlm.nih.gov/24718657/)]
12. Atkeson LR, Adams AN, Alvarez RM. Nonresponse and Mode Effects in Self- and Interviewer-Administered Surveys. *Political Analysis* 2014 May 28;22(3):304-320. [doi: [10.1093/pan/mpt049](https://doi.org/10.1093/pan/mpt049)]
13. Shin E, Johnson TP, Rao K. Survey Mode Effects on Data Quality: Comparison of Web and Mail Modes in a U.S. National Panel Survey. *Social Science Computer Review* 2012 May;30(2):212-228. [doi: [10.1177/0894439311404508](https://doi.org/10.1177/0894439311404508)]
14. Houle B, Angotti N, Clark SJ, Williams J, Gómez-Olivé FX, Menken J, et al. Lets Talk about Sex, Maybe: Interviewers, Respondents, and Sexual Behavior Reporting in Rural South Africa. *Field Methods* 2016 May;28(2):112-132. [doi: [10.1177/1525822X15595343](https://doi.org/10.1177/1525822X15595343)]
15. Olson K, Smyth JD, Wood HM. Does Giving People Their Preferred Survey Mode Actually Increase Survey Participation Rates? An Experimental Examination. *Public Opinion Quarterly* 2012 Aug 02;76(4):611-635. [doi: [10.1093/poq/nfs024](https://doi.org/10.1093/poq/nfs024)]
16. Beebe TJ, McAlpine DD, Ziegenfuss JY, Jenkins S, Haas L, Davern ME. Deployment of a mixed-mode data collection strategy does not reduce nonresponse bias in a general population health survey. *Health Serv Res* 2012 Aug;47(4):1739-1754 [FREE Full text] [doi: [10.1111/j.1475-6773.2011.01369.x](https://doi.org/10.1111/j.1475-6773.2011.01369.x)] [Medline: [22250782](https://pubmed.ncbi.nlm.nih.gov/22250782/)]
17. Chang L, Krosnick JA. National Surveys Via RDD Telephone Interviewing Versus the Internet: Comparing Sample Representativeness and Response Quality. *Public Opinion Quarterly* 2009 Dec 01;73(4):641-678. [doi: [10.1093/poq/nfp075](https://doi.org/10.1093/poq/nfp075)]
18. Tourangeau R, Conrad FG, Couper MP. Measurement Error on the Web and in Other Modes of Data Collection. In: *The Science of Web Surveys*. New York: Oxford University Press; May 09, 2013.
19. Dillman D, Reips U, Matzat U. Advice in Surveying the General Public Over the Internet. *Int J Internet Sci* 2010;5(1):1-4.
20. de Leeuw ED. To Mix or Not to Mix Data Collection Modes in Surveys. *J Off Stat* 2005;21(2):233-255.
21. Burkill S, Copas A, Couper MP, Clifton S, Prah P, Datta J, et al. Using the Web to Collect Data on Sensitive Behaviours: A Study Looking at Mode Effects on the British National Survey of Sexual Attitudes and Lifestyles. *PLoS One* 2016 Feb 11;11(2):e0147983 [FREE Full text] [doi: [10.1371/journal.pone.0147983](https://doi.org/10.1371/journal.pone.0147983)] [Medline: [26866687](https://pubmed.ncbi.nlm.nih.gov/26866687/)]
22. Fricker S, Galesic M, Tourangeau R, Yan T. An Experimental Comparison of Web and Telephone Surveys. *Public Opinion Quarterly* 2005;69(3):370-392. [doi: [10.1093/poq/nfi027](https://doi.org/10.1093/poq/nfi027)]
23. Gosling SD, Vazire S, Srivastava S, John OP. Should we trust web-based studies? A comparative analysis of six preconceptions about internet questionnaires. *Am Psychol* 2004;59(2):93-104. [doi: [10.1037/0003-066X.59.2.93](https://doi.org/10.1037/0003-066X.59.2.93)] [Medline: [14992636](https://pubmed.ncbi.nlm.nih.gov/14992636/)]

24. Poder TG, Bellemare CA, Bédard SK, Lemieux R. Social acceptance and population confidence in telehealth in Quebec. *BMC Health Serv Res* 2015;15:72 [FREE Full text] [doi: [10.1186/s12913-015-0727-1](https://doi.org/10.1186/s12913-015-0727-1)] [Medline: [25889230](https://pubmed.ncbi.nlm.nih.gov/25889230/)]
25. Ansolabehere S, Schaffner BF. Does Survey Mode Still Matter? Findings from a 2010 Multi-Mode Comparison. *Political Analysis* 2014 May 28;22(3):285-303. [doi: [10.1093/pan/mpt025](https://doi.org/10.1093/pan/mpt025)]
26. Livingston MD, Komro KA, Wagenaar AC. The effects of survey modality on adolescents' responses to alcohol use items. *Alcohol Clin Exp Res* 2015 Apr;39(4):710-715. [doi: [10.1111/acer.12659](https://doi.org/10.1111/acer.12659)] [Medline: [25752744](https://pubmed.ncbi.nlm.nih.gov/25752744/)]
27. Plante C, Jacques L, Chevalier S, Fournier M. Comparability of Internet and telephone data in a survey on the respiratory health of children. *Can Respir J* 2012;19(1):13-18 [FREE Full text] [Medline: [22332126](https://pubmed.ncbi.nlm.nih.gov/22332126/)]
28. Link MW, Mokdad AH. Effects of survey mode on self-reports of adult alcohol consumption: a comparison of mail, web and telephone approaches. *J Stud Alcohol* 2005 Mar;66(2):239-245. [Medline: [15957675](https://pubmed.ncbi.nlm.nih.gov/15957675/)]
29. EKOS Research Associates Inc. 2012 Oct. 2012 HIV/AIDS attitudinal tracking survey: Final report URL: <http://www.catie.ca/sites/default/files/2012-HIV-AIDS-attitudinal-tracking-survey-final-report.pdf> [accessed 2015-09-30] [WebCite Cache ID [6bwLZAu21](https://www.webcitation.org/cache/6bwLZAu21)]
30. CANFAR.: The CIHR Social Research Centre in HIV Prevention; 2012. HIV and AIDS in Canada: A National Survey: Summary Report URL: [http://www.srchiv.ca/NationalSurvey/wp-content/uploads/2012/04/Attitudinal-Survey-Summary-Report\\_June15.pdf](http://www.srchiv.ca/NationalSurvey/wp-content/uploads/2012/04/Attitudinal-Survey-Summary-Report_June15.pdf) [accessed 2015-09-30] [WebCite Cache ID [6bwLJxscN](https://www.webcitation.org/cache/6bwLJxscN)]
31. Allman D, Calzavara L, Worthington C, Tyndall M, Adrien A, Walters M, et al. Charitable giving for HIV and AIDS: results from a Canadian national survey. *PLoS One* 2014;9(8):e103184 [FREE Full text] [doi: [10.1371/journal.pone.0103184](https://doi.org/10.1371/journal.pone.0103184)] [Medline: [25153827](https://pubmed.ncbi.nlm.nih.gov/25153827/)]
32. Worthington CA, Calzavara LM, White SJ, Allman D, Tyndall MW. Individual and jurisdictional factors associated with voluntary HIV testing in Canada: Results of a national survey, 2011. *Can J Public Health* 2014 Nov;106(2):e4-e9. [Medline: [25955665](https://pubmed.ncbi.nlm.nih.gov/25955665/)]
33. Austin PC. An Introduction to Propensity Score Methods for Reducing the Effects of Confounding in Observational Studies. *Multivariate Behav Res* 2011 May;46(3):399-424 [FREE Full text] [doi: [10.1080/00273171.2011.568786](https://doi.org/10.1080/00273171.2011.568786)] [Medline: [21818162](https://pubmed.ncbi.nlm.nih.gov/21818162/)]
34. Archer K, Lemeshow S. Goodness-of-fit test for a logistic regression model fitted using survey sample data. *Stata Journal* 2006 Feb;6(1):97-105.
35. Statcan. Ottawa, ON: Statistics Canada; 2011. The Canadian Population in 2011: Age and Sex, 2011 Census URL: <http://www12.statcan.gc.ca/census-recensement/2011/as-sa/98-311-x/98-311-x2011001-eng.pdf> [accessed 2016-07-21] [WebCite Cache ID [6j9soJh2L](https://www.webcitation.org/cache/6j9soJh2L)]
36. Statcan. Ottawa, ON: Statistics Canada; 2012. The Canadian Population in 2011: Population Counts and Growth, 2011 Census URL: <http://www12.statcan.gc.ca/census-recensement/2011/as-sa/98-310-x/98-310-x2011001-eng.pdf> [accessed 2016-07-21] [WebCite Cache ID [6j9svlWHN](https://www.webcitation.org/cache/6j9svlWHN)]
37. Statcan. Ottawa, ON: Statistics Canada; 2013. Education in Canada: Attainment, Field of Study and Location of Study, National Household Survey, 2011 URL: <http://www12.statcan.gc.ca/nhs-enm/2011/as-sa/99-012-x/99-012-x2011001-eng.pdf> [accessed 2016-07-21] [WebCite Cache ID [6j9t8Deri](https://www.webcitation.org/cache/6j9t8Deri)]
38. Poder TG, He J, Simard C, Pasquier JC. Willingness to pay for ovulation induction treatment in case of WHO II anovulation: a study using the contingent valuation method. *Patient Prefer Adherence* 2014;8:1337-1346 [FREE Full text] [doi: [10.2147/PPA.S67742](https://doi.org/10.2147/PPA.S67742)] [Medline: [25328385](https://pubmed.ncbi.nlm.nih.gov/25328385/)]
39. Atkinson NL, Saperstein SL, Pleis J. Using the internet for health-related activities: findings from a national probability sample. *J Med Internet Res* 2009;11(1):e4 [FREE Full text] [doi: [10.2196/jmir.1035](https://doi.org/10.2196/jmir.1035)] [Medline: [19275980](https://pubmed.ncbi.nlm.nih.gov/19275980/)]
40. Martin S, Robinson J. The Income Digital Divide: Trends and Predictions for Levels of Internet Use. *Social Problems* 2007 Feb;54(1):1-22. [doi: [10.1525/sp.2007.54.1.1.HeinOnline](https://doi.org/10.1525/sp.2007.54.1.1.HeinOnline)]
41. Fairlie RW. Explaining differences in access to home computers and the Internet: A comparison of Latino groups to other ethnic and racial groups. *Electron Commerce Res* 2007 Dec;7(3):265-291. [doi: [10.1007/s10660-007-9006-5](https://doi.org/10.1007/s10660-007-9006-5)]
42. Howard P, Busch L, Sheets P. Comparing Digital Divides: Internet Access and Social Inequality in Canada and the United States. *Can J Commun*. 2010. (1) p. 109-128 URL: <http://www.cjc-online.ca/index.php/journal/article/view/2192> [WebCite Cache ID [6jIQpvH2M](https://www.webcitation.org/cache/6jIQpvH2M)]
43. Bekkers R, Wiepking P. Accuracy of self-reports on donations to charitable organizations. *Qual Quant* 2010 Jun;45(6):1369-1383. [doi: [10.1007/s11135-010-9341-9](https://doi.org/10.1007/s11135-010-9341-9)]
44. Fricker RD, Schonlau M. Advantages and Disadvantages of Internet Research Surveys: Evidence from the Literature. *Field Methods* 2002 Nov 01;14(4):347-367. [doi: [10.1177/152582202237725](https://doi.org/10.1177/152582202237725)]
45. Amari E, Vandebeek C, Montgomery CJ, Skarsgard E, Ansermino JM. Telephone and web-based pediatric day surgery questionnaires. *Int J Health Care Qual Assur* 2010;23(3):339-351. [doi: [10.1108/09526861011029398](https://doi.org/10.1108/09526861011029398)] [Medline: [20535905](https://pubmed.ncbi.nlm.nih.gov/20535905/)]
46. Schonlau M. Will web surveys ever become part of mainstream research? *J Med Internet Res* 2004 Sep 23;6(3):e31 [FREE Full text] [doi: [10.2196/jmir.6.3.e31](https://doi.org/10.2196/jmir.6.3.e31)] [Medline: [15471757](https://pubmed.ncbi.nlm.nih.gov/15471757/)]
47. Yeager DS, Krosnick JA, Chang L, Javitz HS, Levendusky MS, Simpser A, et al. Comparing the Accuracy of RDD Telephone Surveys and Internet Surveys Conducted with Probability and Non-Probability Samples. *Public Opinion Quarterly* 2011 Oct 05;75(4):709-747. [doi: [10.1093/poq/nfr020](https://doi.org/10.1093/poq/nfr020)]

48. Galesic M, Tourangeau R, Couper MP. Complementing random-digit-dial telephone surveys with other approaches to collecting sensitive data. *Am J Prev Med* 2006 Nov;31(5):437-443. [doi: [10.1016/j.amepre.2006.07.023](https://doi.org/10.1016/j.amepre.2006.07.023)] [Medline: [17046416](https://pubmed.ncbi.nlm.nih.gov/17046416/)]

## Abbreviations

**AIDS:** acquired immune deficiency syndrome

**CI:** confidence interval

**DK:** do not know

**HIV:** human immunodeficiency virus

**OR:** odds ratio

**SE:** standard error

*Edited by G Eysenbach, P Sullivan, T Sanchez; submitted 30.09.15; peer-reviewed by J Krosnick, T Poder, C Pierannunzi, M Van Handel; comments to author 22.11.15; revised version received 03.01.16; accepted 06.07.16; published 29.07.16.*

*Please cite as:*

*Jones MK, Calzavara L, Allman D, Worthington CA, Tyndall M, Iveniuk J*

*A Comparison of Web and Telephone Responses From a National HIV and AIDS Survey*

*JMIR Public Health Surveill* 2016;2(2):e37

URL: <http://publichealth.jmir.org/2016/2/e37/>

doi: [10.2196/publichealth.5184](https://doi.org/10.2196/publichealth.5184)

PMID: [27473597](https://pubmed.ncbi.nlm.nih.gov/27473597/)

©Marcella K Jones, Liviana Calzavara, Dan Allman, Catherine A Worthington, Mark Tyndall, James Iveniuk. Originally published in *JMIR Public Health and Surveillance* (<http://publichealth.jmir.org>), 29.07.2016. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in *JMIR Public Health and Surveillance*, is properly cited. The complete bibliographic information, a link to the original publication on <http://publichealth.jmir.org>, as well as this copyright and license information must be included.

Original Paper

# Acceptability of a Community-Based Outreach HIV-Testing Intervention Using Oral Fluid Collection Devices and Web-Based HIV Test Result Collection Among Sub-Saharan African Migrants: A Mixed-Method Study

Jasna Loos<sup>1</sup>, MA; Lazare Manirankunda<sup>1</sup>, MD, MSc; Tom Platteau<sup>2</sup>, MSc; Laura Albers<sup>2</sup>, BN; Katrien Fransen<sup>3</sup>, MSc; Tine Vermoesen<sup>3</sup>, BSc; Fiona Namanya<sup>1</sup>, MA; Christiana Nöstlinger<sup>1</sup>, PhD

<sup>1</sup>HIV and Sexual Health Unit, Department of Public Health, Institute of Tropical Medicine, Antwerp, Belgium

<sup>2</sup>Helpcenter-ITM, Department of Clinical Sciences, Institute of Tropical Medicine, Antwerp, Belgium

<sup>3</sup>AIDS Reference Laboratory, Department of Clinical Sciences, Institute of Tropical Medicine, Antwerp, Belgium

**Corresponding Author:**

Jasna Loos, MA

HIV and Sexual Health Unit

Department of Public Health

Institute of Tropical Medicine

Nationalestraat 155

Antwerp, 2000

Belgium

Phone: 32 32470718

Fax: 32 3 247 65 32

Email: [jloos@itg.be](mailto:jloos@itg.be)

## Abstract

**Background:** Late human immunodeficiency virus (HIV) diagnosis is common among sub-Saharan African migrants. To address their barriers to HIV testing uptake and improve timely HIV diagnoses and linkage to care, the outreach HIV testing intervention, “swab2know,” was developed. It combined a community-based approach with innovative testing methods: oral fluid self-sampling and the choice between Web-based HIV test result collections using a secured website or post-test counseling at a sexual health clinic. The sessions included an informational speech delivered by a physician of sub-Saharan African origin and testimonies by community members living with HIV.

**Objectives:** The objectives of this study were to evaluate the intervention’s acceptability among sub-Saharan African migrants and its potential to reach subgroups at higher risk for HIV infection and to identify facilitators and barriers for HIV testing uptake.

**Methods:** This mixed-method study combined qualitative (participant observations and informal interviews with testers and nontesters) and quantitative data (paper–pencil survey, laboratory data, and result collection files). Data were analyzed using a content analytical approach for qualitative and univariate analysis for quantitative data.

**Results:** A total of 10 testing sessions were organized in sub-Saharan African migrant community venues in the city of Antwerp, Belgium, between December 2012 and June 2013. Overall, 18.2% of all people present (N=780) underwent HIV testing; 29.8% of them tested for HIV for the first time, 22.3% did not have a general practitioner, and 21.5% reported 2 or more sexual partners (last 3 months). Overall, 56.3% of participants chose to collect their HIV test results via the protected website. In total, 78.9% collected their results. The qualitative analysis of 137 participant observation field notes showed that personal needs and Internet literacy determined the choice of result collection method. Generally, the oral fluid collection devices were well accepted mainly because sub-Saharan African migrants dislike blood taking. For some participants, the method raised concerns about HIV transmission via saliva. The combination of information sessions, testimonies, and oral fluid collection devices was perceived as effectively reducing thresholds to participation. Acceptability of the intervention differed between individual participants and settings. Acceptance was higher among women, in churches and settings where community leaders were engaged in HIV awareness raising. Higher preventive outcomes were observed in settings with lower acceptance. The presence of the intervention team visualized the magnitude of the HIV epidemic to the public and promoted HIV testing uptake at large, for example, those who declined indicated they would take up testing later.



**Conclusions:** When accompanied by tailored provision of information, outreach HIV testing interventions adopting a community-based approach and innovative methods such as Web-based result collection and oral fluid collection devices are acceptable and reduce thresholds for HIV testing uptake. The swab2know intervention was able to reach sub-Saharan African migrants at risk of HIV infection, and with limited access to regular HIV testing. Among nontesters, the intervention contributed to awareness raising and therefore has a place in a multipronged HIV test promotion strategy.

(*JMIR Public Health Surveill* 2016;2(2):e33) doi:[10.2196/publichealth.5519](https://doi.org/10.2196/publichealth.5519)

## KEYWORDS

oral fluid collection devices; Web-based HIV test result; sub-Saharan African migrants; outreach HIV testing; acceptability

## Introduction

The human immunodeficiency virus (HIV) epidemic among sub-Saharan African migrants in Europe reflects the generalized epidemics in sub-Saharan Africa. Next to men who have sex with men (MSM), HIV is concentrated in the small communities of sub-Saharan African migrants [1], constituting the second largest group affected by HIV in Belgium [2]. Most of them are believed to have acquired HIV heterosexually in their country of origin, but there is increasing evidence to show that post-migration HIV infection is occurring in Europe [3,4]. Structural factors related to migration [5,6], closely knitted sexual networks of sub-Saharan African migrants [7], and high proportions of late diagnoses [1,2] increase sub-Saharan African migrants' risk for HIV infection in host countries. Greater investment in uptake of HIV testing and linkage to care can alter this trend. Early diagnoses reduce the individual morbidity and mortality of patients and health care costs. Although the advantages of HIV testing are well known by the sub-Saharan African migrant communities, they do not translate in regular uptake [8]. A body of literature describes the multiple and intertwined barriers at individual, community, and health care level for this key population [5,8-14]. To reduce thresholds for HIV testing uptake, interventions both in the communities and at health care level [15,16] have been set up in Flanders. Community-level interventions were developed using a participatory approach and adopted outreach and community mobilization strategies to reach the heterogeneous sub-Saharan African migrant communities [15]. Evidence from Africa suggests that outreach HIV testing interventions can be effective in accessing "hard-to-reach" populations [17,18] and lead to earlier diagnosis of HIV [19,20]. A first pilot project, offering free HIV tests in sub-Saharan African migrant and MSM community settings, yielded good primary prevention benefits, but groups with higher risk of undiagnosed HIV infection were not reached [21]. A follow-up intervention, labeled "swab2know," aimed at further reducing thresholds for testing while preserving the preventive outcomes by adopting new, less invasive HIV-testing technologies. To overcome fears and barriers related to blood taking [22], oral fluid collection devices were introduced (Oracol device; Malvern Medical Developments, Worcester, UK). This method was previously validated for the purpose of screening among populations with HIV prevalence above 1%, indicating that it is suitable for sub-Saharan African migrants [23,24]. Result collection often has been described as low in mobile Voluntary HIV counseling and testing (VCT) interventions [21,25]. To reduce failures to return for HIV test results, the option to collect them using a

secured website was given. In addition, result communication through traditional post-test counseling at a low-threshold HIV or sexually transmitted infections (STI) testing facility focused on groups with a high risk of acquiring HIV or STIs like sub-Saharan African migrants were also offered. In this context, low-threshold HIV or STI testing facility is defined as an outpatient testing site offering free and anonymous HIV testing for the main groups at risk for HIV/STIs, that is, MSM, sub-Saharan African migrants, and young people. The post-test counseling approach used was in line with the recommendations made by COBATEST, a European network on community-based voluntary counseling and testing centers, which has acquired extensive experience on HIV community-based testing practices in Europe and published guidance on this topic [26]. Both methodologies—oral fluid sample collection and Web-based test result communication—were novel to Belgium's sub-Saharan African migrant communities. The detailed methodology of the "swab2know" intervention and first results pertaining to testing uptake at MSM community settings in Antwerp city have been described elsewhere [27].

To evaluate this intervention's potential to detect new HIV infections among sub-Saharan African migrants and link them to care (eg, HIV test result collection and referral to HIV-specialized care for people with an HIV diagnosis), we conducted a mixed-method study in different community settings. We describe the key findings in terms of the interventions' acceptability among sub-Saharan African migrants, as well as facilitators and barriers for intervention uptake.

## Methods

### Intervention Methods: "Swab2know" Methodology

Between December 2012 and June 2013, 10 "swab2know" outreach VCT sessions were organized in self-selected community settings of the largest sub-Saharan African migrant communities residing in the city of Antwerp, that is, the Democratic Republic of Congo (DRC), Nigeria, Ghana, and Cameroon. Community settings included churches, bars, events, and information sessions of African organizations. More specifically, we visited a Cameroonian, Nigerian, and Ghanaian organization, an African Lesbian Gay Bisexual Transsexuals (LGBT), an African women organization, and a Congolese youth group. The intervention adopted a settings approach, for example, all people present in the community settings, sub-Saharan African migrants as well as people from other origins were invited to participate, except minors. The rationale behind this approach was that not only sub-Saharan African



migrants are at risk for HIV but also their sexual partners regardless of their origin or migration status.

The “swab2know” sessions included an introduction of the testing offer by a medical doctor of sub-Saharan African origin and a testimony of a community member living with HIV. Interested participants received detailed explanation of the study procedures in a separate area. Participants who accepted to participate were asked to read and sign an informed consent form and to fill a short questionnaire. They self-collected an oral fluid sample through an OraCol collection device. Participants preferring to receive their HIV test result at the HIV or STI testing facility immediately got an appointment. Those choosing the secured website [27] were assisted in creating a personal account through which they could access their results. They received automated email reminders when their results were available. Before the study, community leaders had recommended to not communicate test results indicating an HIV diagnosis via the website. Following this advice, participants were informed that in case of reactive or invalid samples, they would receive a mobile phone call from the study nurse. All data items were linked with a unique code to guarantee confidentiality in handling the data. Within 7 days of collecting the sample, an HIV ELISA (Genscreen HIV-1/2

version 2 BioRad, Marnes-la-Coquette, France) was performed at the AIDS Reference Laboratory of the Institute of Tropical Medicine, according to a validated protocol for oral fluid specimens. This test gives a 100% (95% CI: 95.9-100) sensitivity and 97.6% (95% CI: 94.5-99.0) specificity [28] in established HIV-untreated individuals. In addition, the quality of the oral fluid samples was measured using the IgG ELISA quantification kit (Human IgG ELISA Immunology Consultants Laboratory, Inc., Portland, OR, USA). If the sample contained more than 3500 ng total IgG/mL, it was considered valid.

Approximately 1 week after participation, HIV test results were made available to the participants. As recommended by the community leaders, participants with a reactive HIV result received a mobile phone call from the study nurse inviting them for counseling at the STI and HIV testing center. During the posttest counseling session, it was explained that results indicative for an HIV infection required confirmation by a traditional blood test, which was strongly advised. Participants whose results were confirmed were referred to the AIDS Reference Clinic of the Institute of Tropical Medicine. Participants with a nonreactive sample who chose to collect their result via the secured website found the message as shown in [Textbox 1](#).

**Textbox 1.** Website message delivered to participants with a nonreactive HIV test result for Web-based test result collection (available in Dutch, French, and English).

Your saliva has not reacted to the test. We can conclude with high reliability that you are not infected with HIV. Please note that an HIV test does not give any information about the risk you may have had during the last three months. Furthermore, an oral fluid test is not officially recognized to diagnose an HIV infection.

If you ran a sexual risk during the last three months it is better to have a combined HIV test performed on your blood (antibody and antigen determination). This will give a better (but not absolute) result for recent risks. After having had unsafe sexual contacts, it can be important to test for sexually transmitted infections. To discuss the necessity of such tests, you can contact your GP, an HIV specialized treatment center or Helpcenter. ([www.helpcenteritg.be](http://www.helpcenteritg.be))

When you have casual sex with different partners, it is recommended to have an HIV test every 3-6 months.

We 'd like to contact you by email or text message in 3-6 months, so you can order a test kit for a new oral fluid sample.

If you don't want to receive this email, please let us know by changing the settings of your account.

If we send a test to you, you can collect the sample yourself and send it to ITM's laboratory using the envelope provided. The procedure for obtaining your result will be the same.

In case participants missed the initial appointment or failed to pick up the results from the website, the study nurse made up to 4 attempts to contact participants and encourage them to collect their results to ensure linkage to care. In case of reactive test results, up to 8 contact attempts were made.

## Research Methods

This prospective interventional study used mixed methods to answer the research questions, capitalizing on each method's strengths to obtain a nuanced understanding of the intervention's potential to detect new HIV infections and to link participants to care.

## Quantitative Methods

To identify participants' characteristics, including their risk for an undiagnosed HIV infection, a brief self-reported 11-item questionnaire was used to assess sociodemographics, HIV testing and sexual behavior, and access to health care. Questionnaire data were linked to laboratory data and result

collection records to identify facilitators of result collection and preferred communication method.

Quantitative data were analyzed using SPSS 22 software (IBM). A first descriptive analysis was performed on all variables and stratified by gender. In the second univariate analysis, Pearson chi-square was used according to variable properties to compare proportions and identify factors associated with preference in result collection methods and actual result collection. A significance level of 5% was applied.

## Qualitative Methods

For an in-depth assessment of acceptability, facilitators, and barriers for participation, self-reported data were complemented with qualitative data from participant observations during 10 testing events. Two trained female social scientists, of Ugandan and Belgian origin, observed the 10 sessions and conducted informal interviews in English, French, or Dutch with 3 groups: testers, nontesters, and members of the intervention team. The

researchers mingled with the crowd during the informational speech to observe people's verbal and nonverbal reactions. Through this approach, people with particular viewpoints were identified and approached for an informal interview [29] striving for maximum variation in the sample. We adopted an unstructured interview approach [30], probing for the following themes: (1) perception and acceptability of outreach VCT in sub-Saharan African migrant community settings, (2) motivations for declining or accepting the test offer, (3) perceptions of and/or experiences with the oral fluid collection devices, (4) perception and acceptability of HIV test result collection via a secured website or a low threshold HIV and STI testing center. During participant observation and after informal interviews, the researchers took jottings, which were assembled into field notes the day after the session. Field notes were uploaded in N-VIVO 10, coded using a data-driven codebook established by 2 independent researchers. Following inductive analysis principles [31], emerging themes and their relationships were identified. Text segments that were coded differently were discussed until consensus was reached. A second-order analysis identified participant clusters based on gender, age, nationality, and type of setting.

### **Ethical Considerations**

The swab2know intervention was conceptualized and implemented in close collaboration with sub-Saharan African migrant community leaders. Ethical approval was obtained from the Institutional Review Board at the Institute of Tropical Medicine and the University Hospital in Antwerp.

## **Results**

### **Quantitative Findings**

#### ***Characteristics, Settings, and Participants***

About 780 people were present in the selected venues and participated in the information session, and 142 (18.2%) underwent HIV testing. [Table 1](#) summarizes participants' characteristics and HIV test results.

#### ***HIV Testing Results and Behavioral Characteristics***

The samples of 5 participants, all from men, were indicative of an HIV infection. About 29.8% of participants, more men than women, tested for the first time. An additional 44.1% reported that their last HIV test dated from more than 1 year ago, and 46.2% did not know their current HIV status, 22.3% had no GP, and 21.5% reported to have had 2 or more sexual partners during the last 3 months.

#### ***Result Collection and Linkage to Care***

Of all participants, 56.3% chose to collect their results via the secured website, 43.7% opted for a consultation at the HIV and STI testing center. Participants who decided for Web-based result collection were more likely able to speak Dutch ( $P=.003$ ) and came from settings with lower general acceptance of the intervention, as identified based on the qualitative data (see "Acceptability of Outreach HIV Testing in African Migrant Community Settings" section). We observed that at events and in bars, participants were more likely to opt for Web-based result collection, but this could not be confirmed statistically. The same was found for participants who already had tested before and those living abroad.

Overall, 78.9% of participants collected their HIV test result. Of those who opted for result collection at the testing facility, 85% collected their HIV test result, compared with 74% who chose the website. This difference was statistically not significant. Five participants who initially preferred to collect their results through the Internet came to the testing center because they failed to access the website. None of the associations with the variable result collection tested in the univariate analysis (as shown in [Table 2](#)) were statistically significant. Of the 5 participants with a reactive test result, 2 were confirmed HIV positive on blood, of which 1 was a new diagnosis and 1 was a patient already in medical follow-up. One test result was confirmed to be false positive on a blood sample. Two participants with a reactive result never showed up for confirmation testing. They were contacted by the study nurse and invited for counseling at the HIV and STI testing center. When they missed the first appointment, 8 additional attempts were made to contact them again, without success.

**Table 1.** Participants' characteristics and HIV test results, stratified by gender.

	Total <sup>a</sup> N:142 (%)	Male N:73 (%)	Female N:60 (%)
<b>Setting type</b>			
Church	32 (22.5%)	12 (16%)	15 (25%)
Event	42 (29.6%)	29 (40%)	13 (22%)
Bar	14 (9.9%)	12 (16%)	0 (0%)
Information session	53 (37.3%)	20 (27%)	32 (53%)
<b>Survey language</b>			
English	68 (47.9%)	29 (40%)	33 (55%)
French	50 (35.2%)	33 (45%)	14 (23%)
Dutch	24 (16.9%)	11 (22%)	13 (22%)
<b>Age, years</b>			
Mean (rang)	38.5 (17-73)	38.2 (17-60)	38.6 (18-73)
≤ 30	41 (29.3%)	20 (27%)	20 (33%)
31-40	41 (29.3%)	27 (37%)	12 (20%)
41-50	30 (21.4%)	11 (15%)	17 (28%)
≥ 51	28 (20.0%)	15 (21%)	11 (18%)
<b>Country of origin</b>			
Central Africa (DRC, Angola, Cameroon)	61 (43.0%)	40 (55%)	17 (28%)
East Africa (Burundi, Kenya, Somalia, Tanzania, Uganda, Rwanda, and Zambia)	10 (7.0%)	4 (5%)	6 (10%)
West Africa (Ivory coast, Ghana, Mali, Liberia, Nigeria, Senegal, Cape Verde)	51 (35.9%)	24 (33%)	22 (37%)
Southern Africa (South Africa)	2 (1.4%)	0 (0%)	2 (3%)
European (Belgium and France)	8 (5.6%)	3 (4%)	5 (8%)
Other (Curaçao, Jamaica, Suriname, Guyana, and Morocco)	10 (7.0%)	2 (3%)	8 (13%)
<b>Country of residence</b>			
Belgium	135 (95.1%)	70 (96%)	56 (93%)
Other (Netherlands, UK, France, Sweden, and Canada)	7 (4.9%)	3 (4%)	4 (7%)
<b>Access to primary care<sup>b</sup></b>			
Has a general practitioner (GP)	108 (77.7%)	53 (76%)	48 (80%)
Does not have a GP	31 (22.3%)	17 (24%)	12 (20%)
<b>Sexual preference<sup>c</sup></b>			
Heterosexual	109 (93.2%)	59 (92%)	50 (94%)
Gay, Lesbian, or bisexual	8 (6.8%)	5 (8%)	3 (6%)
<b>Number of sexual partners in last 3 months<sup>d</sup></b>			
None	29 (21.5%)	12 (17%)	13 (22%)
1	77 (57.0%)	40 (57%)	35 (59%)
2-6	29 (21.5%)	18 (26%)	11 (19%)
<b>Last HIV test<sup>e</sup></b>			
Less than 1 year ago	35 (26.1%)	21 (31%)	12 (21%)

	Total <sup>a</sup> N:142 (%)	Male N:73 (%)	Female N:60 (%)
Between 1 and 3 years ago	36 (26.9%)	15 (22%)	19 (33%)
More than 3 years ago	23 (17.2%)	10 (15%)	12 (21%)
Never tested	40 (29.8%)	22 (32%)	14 (25%)
Reported believed HIV status (questionnaire data) <sup>f</sup>			
HIV-negative	70 (53.8%)	34 (52%)	33 (57%)
Unknown	60 (46.2%)	32 (48%)	25 (43%)
Result oral fluid sample (laboratory data)			
HIV-negative	137 (96.5%)	68 (93%)	60 (100%)
Reactive	5 (3.5%)	5 (7%)	0 (0%)

<sup>a</sup>For 9 participants, gender data are missing.

<sup>b</sup>Data missing of 3 participants.

<sup>c</sup>Data missing of 25 participants.

<sup>d</sup>Data missing of 7 participants.

<sup>e</sup>Data missing of 8 participants.

<sup>f</sup>Data missing of 12 participants.

**Table 2.** Comparison of participant and intervention characteristics for preferred HIV test result collection method and result collection (N=142).

	Collection method		<i>P</i> value	HIV test result collected		<i>P</i> value
	Testing center	Secured website		Yes	No	
	N:62 (43.7%)	N:80 (56.3%)		N:112 (78.9%)	N:30 (21.1%)	
Setting type						
Church	16 (50%)	16 (50%)	.06	28 (87.5%)	4 (12.5%)	.20
Event or bar	18 (32%)	39 (68%)		41 (72%)	16 (28%)	
Information session	28 (53%)	25 (47%)		43 (81%)	10 (19%)	
Acceptability in setting <sup>a</sup>						
High acceptance	42 (52%)	39 (48%)	.02 <sup>c</sup>	67 (83%)	14 (17%)	.20
Low acceptance	20 (33%)	41 (67%)		45 (74%)	16 (26%)	
Survey language						
English	34 (50%)	34 (50%)	.003 <sup>c</sup>	56 (82%)	12 (18%)	.26
French	25 (50%)	25 (50%)		40 (80%)	10 (20%)	
Dutch	3 (12.5%)	21 (87.5%)		16 (67%)	8 (33%)	
Result collection method						
HIV and STI testing center			<sup>d</sup>	53 (85%)	9 (15%)	.09
Secured website				59 (74%)	21 (26%)	
Gender <sup>b</sup>						
Male	37 (51%)	36 (49%)	.10	56 (77%)	17 (23%)	.65
Female	22 (37%)	38 (63%)		48 (80%)	12 (20%)	
Age <sup>b</sup> , years						
≤ 30	13 (32%)	28 (68%)	.17	27 (66%)	14 (34%)	.06
31-40	23 (56%)	18 (44%)		37 (90%)	4 (10%)	
41-50	13 (43%)	17 (57%)		24 (80%)	6 (20%)	
≥ 51	13 (46%)	15 (54%)		22 (79%)	6 (21%)	
Country of origin						
Central Africa	25 (41%)	36 (59%)	<sup>d</sup>	45 (74%)	16 (26%)	<sup>d</sup>
East Africa	4 (40%)	6 (60%)		7 (70%)	3 (30%)	
West Africa	27 (53%)	24 (47%)		43 (84%)	8 (16%)	
Southern Africa	0 (0%)	2 (100%)		2 (100%)	0 (0%)	
Belgium and France	3 (37.5%)	5 (62.5%)		7 (87.5%)	1 (12.5%)	
Other	3 (20%)	7 (70%)		8 (80%)	2 (20%)	
Residence						
Belgium	62 (45.9%)	73 (54.1%)	<sup>d</sup>	106 (78.5%)	29 (21.5%)	<sup>d</sup>
Abroad	0 (0%)	7 (100%)		6 (86%)	1 (14%)	
General practitioner <sup>b</sup>						
No	13 (42%)	18 (58%)	.88	25 (81%)	6 (19%)	.82
Yes	47 (43.5%)	61 (56.5%)		85 (78.7%)	23 (21.3%)	
Sexual Preference <sup>b</sup>						
Heterosexual	50 (46.7%)	57 (53.3%)	<sup>d</sup>	85 (80.1%)	21 (19.8%)	<sup>d</sup>
Gay, Lesbian, or bisexual	5 (62.5%)	3 (37.5%)		5 (62.5%)	3 (37.5%)	
Number of partners <sup>b</sup>						



	Collection method		<i>P</i> value	HIV test result collected		<i>P</i> value
	Testing center	Secured website		Yes	No	
	N:62 (43.7%)	N:80 (56.3%)		N:112 (78.9%)	N:30 (21.1%)	
None	14 (48%)	15 (52%)	.64	20 (69%)	9 (31%)	.08
1	33 (43%)	43 (57%)		60 (79%)	16 (21%)	
2-6	15 (54%)	13 (46%)		26 (93%)	2 (7%)	
Testing history						
Never tested	22 (55%)	18 (45%)	.08	31 (77.5%)	9 (22.5%)	.82
Ever tested	39 (38.6%)	62 (61.4%)		80 (79.2%)	21 (20.8%)	
Reported HIV status <sup>b</sup>						
HIV negative	26 (37%)	44 (63%)	.27	58 (83%)	12 (17%)	.27
Unknown	28 (47%)	32 (53%)		45 (75%)	15 (25%)	
Result oral fluid sample (laboratory data)						
Negative	60 (46.8%)	77 (53.2%)	<sup>d</sup>	109 (79.6%)	28 (20.4%)	<sup>d</sup>
Reactive HIV test	2 (40%)	3 (60%)		3 (60%)	2 (40%)	

<sup>a</sup>Categorization based on the qualitative data (see “Acceptability of Outreach HIV Testing in African Migrant Community Settings” paragraph).

<sup>b</sup>Cases with missing data have been excluded from the univariate analysis.

<sup>c</sup>Significant difference at 5% level.

<sup>d</sup>Indicates that *P* value could not be calculated because of the limited number of cases in 1 of the variables.

## Qualitative Findings

We collected a total of 137 field notes, 41 of informal interviews from testers who tested for HIV, 53 of nontesters, and 16 of people who were still undecided. In addition, 14 conversations were held with the intervention team, and 13 descriptive notes of observations were collected.

### Acceptability of Outreach HIV Testing in African Migrant Community Settings

Based on the qualitative data, settings with high versus low acceptance for outreach HIV testing interventions emerged. Acceptance referred to people’s attitudes toward the intervention and its methods, that is, participants expressed appreciation, which did not necessarily translate into participation. In settings frequented by women and adults aged between 30 and 50 years, for example, churches and women’s organizations, acceptance was rather high. Other factors related to acceptance clearly emerging from the data were community ownership, prior HIV awareness, and the perceived appropriateness to test in such settings. Moreover, communities familiar with HIV prevention campaigns, such as the Congolese and the LGBT communities, and also those whose when leaders had sensitized their members in advance, accepted the intervention more easily than others. At the World AIDS day event, a man aged 40 to 50 years old told me:

*This event is a good day to test. It is a mixture of celebrations to remember the people who have passed away and those still living with HIV.*

*At a Nigerian cultural event, a man in his forties told me he could not understand why this special event had to be interrupted by an activity that has nothing*

*to do with the festival (...) “The focus should be on the king and nothing else.*

The aforementioned quotes demonstrate the close link between general acceptance and the perceived appropriateness. Mixing HIV with cultural celebrations and parties was disapproved of, particularly when leaders failed to openly take ownership for the testing offer. Settings, where statements like “Go to the people who need it” were commonly heard, for instance, calling on “prostitutes and womanizers in need of HIV-testing,” were subsequently categorized as settings with lower acceptance. In such cases, members of the audience often did not perceive themselves as being at risk of HIV and therefore accepted outreach HIV testing to a lesser extent. Regardless of the confidential procedure, participants seemed to fear social control, due to closely knitted social networks: At the information evening of the Ghanaian association, a lady in her thirties told me:

*Someone can ask: “Oh, last time you went for testing. How was it?” Then you cannot tell lies all the time, you will have to tell them. If you went to the doctor privately, you can choose the time to share such news.*

In settings with lower acceptance, however, prevention benefits seemed to be larger. This finding is supported through the laboratory data, as the 5 reactive cases were all found in the settings with lower acceptance. Many people at these settings also indicated that the information session and in particular, the testimony of a community member living with HIV had raised their awareness of HIV and changed their views about HIV testing. They often mentioned their intention to test later, for instance at their GP or asked concrete information about the HIV and STI testing center.

*At the Nigerian cultural event, I approached a lady in her thirties. When I asked her thoughts about the testing offer, she answers: "Take that message to the people who need it." (...) When the witness came back to her seat [after her testimony], I saw that she asked her for her business card (...) When I approached her again, she explained she would not test at that particular day but when she reaches home she would read more information on the website.*

### **Facilitators and Barriers for Uptake of HIV-Testing**

Participants expressed their motivation for testing often through statements like "it is good to know your status." Participants said the team's presence reminded them of the need to test and they felt encouraged by the informational speech and testimony. Participants also wanted to benefit of the opportunity to get a free test without having to visit a physician.

*At a church, a man in his forties told me that they [the swab2know team] explained it well and it is a good occasion to test and know it. If you go for checkups, like headache or something else, there is no time for a test. He thinks it's good that he doesn't have to go to the hospital and waste more time there.*

Only few participants referred to their previous testing behavior: either they never tested before, tested regularly, or they wanted to confirm previous test results. In some cases, peer and partner pressure were also at play:

*At church, a woman in her thirties told me she was not going to test because she had given birth to all of her children in Belgium and "surely, if there was something wrong the doctors would have said it. (...) A bit later the pastor's wife joined our conversation and told her she was going to have a test. The first lady changed her mind and went along. Later she told me: "it feels good to test, it was good and simple."*

As this quote illustrates, people who declined referred to a previous test. However, probing showed that Africans often assumed they had been tested. Some participants believe that an HIV test is standard procedure with every blood analysis carried out, and because "the doctor never mentioned anything," they were convinced to be HIV negative. Men also believed frequently to be HIV negative by association:

*At the Cameroonian cultural event, a man in his thirties told me he sees no reason to test. (...) All his previous girlfriends have gone to marry and give birth.*

Others declined the testing offer because they assessed their personal risk to be low, for instance, women, who were married or reported to use condoms consistently. Men made risk assessments of their partners:

*Arriving at the Nigerian event, I started talking to a young boy aged 20-30 who is the photographer. He tells me that he has a "non-serious girlfriend" (...) When I ask him what he thinks of the testing offer he says: "I am only with my girlfriend, I don't have to test." When I ask him if he knows her HIV status, he*

*says: "he virginized her" (...) "I am not a womanizer. There are lots of womanizers. Also I, myself, I can get a lot of women with this business, the girls come and show off in front of my camera (...) "No, I never tested, but I am sure I don't have it. I trust my girl."*

### **Perception of Oral Fluid Collection Devices**

Almost all testers said that the oral fluid collection devices were easy to use and painless. They welcomed them as a good alternative for blood samples:

*The Ghanaian pastor in his forties told me that coming to church he was thinking about how they were going to take the blood and "drain us." He thought they were going to take three bottles, a small and two big ones and in the end the result would be the same. He was happy with the "saliva" test, he said "new technology is good, because people don't like blood taking."*

Clearly, the use of oral fluid collection devices had convinced many participants to test and as such lowered their thresholds for HIV testing. However, the method also raised misconceptions pertaining to HIV transmission modes, which needed to be addressed by the intervention team:

*At the Nigerian cultural event a women in her thirties told me that she never would have guessed that now HIV can be found in saliva. "What's next? Are we now going to find it in sweat or if we shake hands?" I told her that she can ask the doctor but HIV could not be transmitted through handshakes. She answered "My sister, don't be too sure, it is also liquid like saliva."*

### **Web-Based Result Collection Versus Counseling at the HIV or STI Testing Center**

The informal interviews also focused on the result collection method chosen and participants' motives for their decision. Those who selected the website appreciated the convenience of not having to comply with the opening hours of the HIV and STI testing center. Some added "the website is okay, because I do not have anything." Others said it was their incentive to test because it enabled participation of sub-Saharan African migrants living outside the Antwerp region.

Participants who chose for result collection via the testing center mostly referred to technical barriers in relation to the website, such as no Internet access, no email address, or low Internet skills. Interestingly, only few expressed the need for professional counseling:

*At church, a man in his twenties told me that he wants his results from Helpcenter-ITG because "with the website you might not be able to understand it well, there might be miscommunication. It is good that they sit you down and explain to you what you have to do."*

## Discussion

This study adopted mixed methods to assess the acceptability of outreach VCT, using oral fluid collection devices and offering Web-based result collection among a sample of sub-Saharan African migrants. Our findings show that when accompanied by tailored provision of information, these innovative methods are acceptable and are perceived to reduce thresholds for HIV testing uptake. Using a community-based participatory settings approach, we were able to reach migrants from high endemic regions, that is, 87.3% originated from sub-Saharan Africa. Almost 1 in 5 participants (ie, 18.2%) present at the respective settings participated. This is considerably more than in a previous intervention undertaken in comparable community settings, where only 8.7% of the people present gave a blood sample for HIV testing, and results were only available from the HIV and STI testing center [21]. The qualitative data showed that especially sampling via oral fluid collection devices reduced barriers to uptake because sub-Saharan African migrants dislike blood taking. Offering the choice between 2 result collection methods—via a secured website or STI or HIV testing facility—attracted participants with different needs. Although only 1 participant was successfully linked to HIV care, the intervention shows some potential to reach subgroups at increased risk for HIV and with reduced access to mainstream testing opportunities. Overall, 29.8% reported testing for HIV for the first time, 22.3% did not have a GP, and 21.5% reported 2 or more sexual partners in the last 3 months.

The mixed-method research approach of quantitative and ethnographic methods to assess acceptability resulted in in-depth knowledge of the facilitating and hindering factors, describing the complexity of the decision-making process for testing uptake. Uptake was highly dependent of leaders' ownership and perceived appropriateness of the event, as shown in similar initiatives [32]. Acceptance was highest in churches, underlining their potential for future outreach testing. Similar to others, we identified that HIV test result collection was better in church settings [33], confirming the influential role of religious leaders in HIV stigma reduction and HIV testing uptake. Although overall, more men than women participated—which can be explained by the male dominance in community settings—the qualitative data revealed that men were less accepting of the testing initiative and more reluctant to test. This is in line with UK research [34,35]. Motivating their refusal, especially men seem to make risk assessments based on assumptions of their current and previous partners and themselves. In addition, misconceptions about how HIV tests would be regularly performed by health care providers and transferability of the (ex-) partners' test result were still prevailing. These qualitative findings have implications for HIV prevention, calling for a gender-specific approach and for addressing culturally grounded misconceptions. Comparable to similar research [36], we identified that sub-Saharan African migrants with poor protective behavior reported low perceived risk. This stresses the importance of qualitative and tailored provision of information accompanying outreach HIV testing.

Most testers in our study were opportunistic testers [26], motivated by their preexisting attitudes toward health and HIV

testing. Those already convinced of the importance of regular testing perceived the noninvasive oral fluid collection devices as an additional incentive. Testimony given by community members living with HIV was able to motivate initially skeptical participants, which underlines the importance of involving people living with HIV in HIV prevention [26,37]. However, the qualitative data also demonstrated the effect of peer and partner pressure on decision taking, which may be typical for close-knitted communities. Compared with another outreach HIV testing project with delayed result communication among African Americans (46% among uninfected and 66% among infected participants; [25]), result collection was high (79%) in our study. Yet, from the perspective of linkage to care, the following considerations need to be made. Participants who opted for the website delivery of the HIV test results more often failed to pick up their results than participants who chose delivery at the testing center. Qualitative data identified lack of technical skills and access to the Internet as main barriers, indicating that communicating results only through a secured website are not yet advisable for this target group. Regardless of our efforts to assure linkage to care, we failed to follow-up 2 of 5 persons with a reactive test results. As late diagnosis and delayed presentation to care after HIV diagnosis are more common in the subgroup of sub-Saharan African migrants [38], decreasing anonymity in favor of collecting more person-identifying data should be debated. To increase participation, we opted to fully anonymize data. However, the qualitative data showed that the spatial limitation of certain settings, closely knitted social networks, and influence of social desirability and leaders undermine these efforts. In practice, anonymous participation can neither be fully guaranteed nor could it be identified as a decisive factor for HIV test uptake. Rather, confidentiality and reliability of the test results are priorities for sub-Saharan African migrants [39]. Indicating that collecting more person-identifying data to improve linkage to care would be acceptable.

This study has some methodological limitations: We relied partly on self-reported data; thus, participants may have given social desirable answers. Study participants were self-selected, which potentially introduced a bias. The qualitative data collection was carried out among a subsample of all intervention participants, and although we have tried to maximize their variation, this may also have influenced the outcome. Due to the practical constraints at the settings, we could not afford to ask detailed questions in the questionnaire. For instance, questionnaire data on sexual behavior did not include past and current preventive behaviors such as condom use. Finally, the study was limited to 10 outreach HIV testing sessions and included a small convenience sample; therefore, results cannot be generalized. Notwithstanding these limitations, the findings generate interesting implications for future community-based outreach HIV testing interventions among migrants.

## Conclusion

This mixed-method evaluation of the “Swab2know” intervention targeting sub-Saharan African migrants yielded promising results. Such interventions need to be replicated and evaluated more rigorously before upscaling. The intervention was able to reach sub-Saharan African migrants and other migrants from



high endemic regions at higher risk of HIV infection and with limited access to regular HIV testing: almost, a third tested for the first time, 21.5% reported having had 2 or more sexual partners in the last 3 months, and 22.3% reported not having a GP. The qualitative data demonstrated the additional prevention benefit of outreach VCT. Prevention outcome was high especially in settings with lower acceptance: for many participants, the presence of the intervention team visualized

the magnitude of the HIV epidemic and promotes HIV testing uptake at large. The fact that many people indicated they would postpone testing to a later moment, for instance, with their GP, calls for a multipronged HIV test promotion strategy targeting communities and health care providers alike [40] to increase HIV testing uptake among vulnerable and diverse migrant communities.

## Acknowledgments

The authors wish to thank all study participants for their contribution to the study and the leaders of the community settings for their engagement in facilitating the intervention. They also acknowledge Veronica van Wijk and Sabien Salomez for help with editing and reference managing.

## Conflicts of Interest

None declared.

## References

1. European Centre for Disease Prevention Control, WHO Regional Office for Europe. HIV/AIDS surveillance in Europe 2012. Stockholm: ECDC; 2013. URL: [http://www.euro.who.int/\\_data/assets/pdf\\_file/0018/235440/e96953.pdf?ua=1;www.webcitation.org/6eSXXUUY](http://www.euro.who.int/_data/assets/pdf_file/0018/235440/e96953.pdf?ua=1;www.webcitation.org/6eSXXUUY) [accessed 2016-07-07] [WebCite Cache ID 6ioyrHbph]
2. Sasse A, Deblonde J, Van Beckhoven D. Epidemiologie van AIDS en HIV infectie in België. Toestand op 31 december 2014. Brussels: WIV-ISP; 2015. URL: [https://www.wiv-isp.be/News/Documents/Rapport\\_HIV-AIDS\\_2013\\_Print\\_Press.pdf](https://www.wiv-isp.be/News/Documents/Rapport_HIV-AIDS_2013_Print_Press.pdf) [WebCite Cache ID 6eSXCqprf]
3. Alvarez-del Arco D, Fakoya I, Monge S, Gennotte AF, Touloumi G, Zuure F, aMASE Study Group. HIV Acquisition among Migrants Living in Europe: Results from aMASE - Advancing Migrant Access to Health Services in Europe. 2015 Presented at: Fifteenth European AIDS Conference; October 21-24, 2015; Barcelona.
4. Fakoya I, Reynolds R, Caswell G, Shiripinda I. Barriers to HIV testing for migrant black Africans in Western Europe. *HIV Med* 2008 Jul;9 Suppl 2:23-25 [FREE Full text] [doi: [10.1111/j.1468-1293.2008.00587.x](https://doi.org/10.1111/j.1468-1293.2008.00587.x)] [Medline: [18557866](https://pubmed.ncbi.nlm.nih.gov/18557866/)]
5. Alvarez-del AD, Monge S, Azcoaga A, Rio I, Hernando V, Gonzalez C, et al. HIV testing and counselling for migrant populations living in high-income countries: a systematic review. *Eur J Public Health* 2013 Dec;23(6):1039-1045 [FREE Full text] [doi: [10.1093/eurpub/cks130](https://doi.org/10.1093/eurpub/cks130)] [Medline: [23002238](https://pubmed.ncbi.nlm.nih.gov/23002238/)]
6. European Centre for Disease Prevention Control. Migrant health: HIV testing and counselling in migrant populations and ethnic minorities in EU/EEA/EFTA Member States. Stockholm: ECDC; 2013. URL: [http://ecdc.europa.eu/en/publications/Publications/1108\\_TER\\_HIV\\_in\\_migrants.pdf](http://ecdc.europa.eu/en/publications/Publications/1108_TER_HIV_in_migrants.pdf) [WebCite Cache ID 6eTgKevPt]
7. Marsicano E, Lydié N, Bajos N. 'Migrants from over there' or 'racial minority here'? Sexual networks and prevention practices among sub-Saharan African migrants in France. *Cult Health Sex* 2013;15(7):819-835. [doi: [10.1080/13691058.2013.785024](https://doi.org/10.1080/13691058.2013.785024)] [Medline: [23659520](https://pubmed.ncbi.nlm.nih.gov/23659520/)]
8. Manirankunda L, Loos J, Alou TA, Colebunders R, Nöstlinger C. "It's better not to know": perceived barriers to HIV voluntary counseling and testing among sub-Saharan African migrants in Belgium. *AIDS Educ Prev* 2009 Dec;21(6):582-593. [doi: [10.1521/aeap.2009.21.6.582](https://doi.org/10.1521/aeap.2009.21.6.582)] [Medline: [20030501](https://pubmed.ncbi.nlm.nih.gov/20030501/)]
9. Adedimeji AA, Asibon A, O'Connor G, Carson R, Cowan E, McKinley P, et al. Increasing HIV testing among African immigrants in Ireland: challenges and opportunities. *J Immigr Minor Health* 2015 Feb;17(1):89-95. [doi: [10.1007/s10903-014-9986-x](https://doi.org/10.1007/s10903-014-9986-x)] [Medline: [24549490](https://pubmed.ncbi.nlm.nih.gov/24549490/)]
10. Burns FM, Johnson AM, Nazroo J, Ainsworth J, Anderson J, Fakoya A, et al. Missed opportunities for earlier HIV diagnosis within primary and secondary healthcare settings in the UK. *AIDS* 2008 Jan 2;22(1):115-122. [doi: [10.1097/QAD.0b013e3282f1d4b6](https://doi.org/10.1097/QAD.0b013e3282f1d4b6)] [Medline: [18090399](https://pubmed.ncbi.nlm.nih.gov/18090399/)]
11. Manirankunda L, Loos J, Debackaere P, Nöstlinger C. "It is not easy": challenges for provider-initiated HIV testing and counseling in Flanders, Belgium. *AIDS Educ Prev* 2012 Oct;24(5):456-468. [doi: [10.1521/aeap.2012.24.5.456](https://doi.org/10.1521/aeap.2012.24.5.456)] [Medline: [23016506](https://pubmed.ncbi.nlm.nih.gov/23016506/)]
12. Ojikutu B, Nnaji C, Sithole J, Schneider KL, Higgins-Biddle M, Cranston K, et al. All black people are not alike: differences in HIV testing patterns, knowledge, and experience of stigma between U.S.-born and non-U.S.-born blacks in Massachusetts. *AIDS Patient Care STDS* 2013 Jan;27(1):45-54 [FREE Full text] [doi: [10.1089/apc.2012.0312](https://doi.org/10.1089/apc.2012.0312)] [Medline: [23259482](https://pubmed.ncbi.nlm.nih.gov/23259482/)]
13. Rice B, Delpech V, Sadler KE, Yin Z, Elford J. HIV testing in black Africans living in England. *Epidemiol Infect* 2013 Aug;141(8):1741-1748. [doi: [10.1017/S095026881200221X](https://doi.org/10.1017/S095026881200221X)] [Medline: [23040613](https://pubmed.ncbi.nlm.nih.gov/23040613/)]
14. Shangase P, Egbe CO. Barriers to accessing HIV services for Black African communities in Cambridgeshire, the United Kingdom. *J Community Health* 2015 Feb;40(1):20-26. [doi: [10.1007/s10900-014-9889-8](https://doi.org/10.1007/s10900-014-9889-8)] [Medline: [24878614](https://pubmed.ncbi.nlm.nih.gov/24878614/)]

15. HIV-SAM Project. Vlaamse organisatie met terreinwerking voor seksuele gezondheid bij Subsaharaanse Afrikaanse migranten. Beleidsplan 2012-2016. Antwerp: Institute of Tropical Medicine; 2011. URL: [https://www.zorg-en-gezondheid.be/sites/default/files/atoms/files/11-4112%20BHO\\_OT\\_SAM\\_2012\\_beleidsplan2012-2016\\_revised.pdf](https://www.zorg-en-gezondheid.be/sites/default/files/atoms/files/11-4112%20BHO_OT_SAM_2012_beleidsplan2012-2016_revised.pdf) [WebCite Cache ID 6eSWBV0hT]
16. Loos J, Manirankunda L, Hendrickx K, Remmen R, Nöstlinger C. HIV testing in primary care: feasibility and acceptability of provider initiated HIV testing and counseling for sub-Saharan African migrants. *AIDS Educ Prev* 2014 Feb;26(1):81-93. [doi: [10.1521/aeap.2014.26.1.81](https://doi.org/10.1521/aeap.2014.26.1.81)] [Medline: [24450280](https://pubmed.ncbi.nlm.nih.gov/24450280/)]
17. Grabbe KL, Menzies N, Taegtmeier M, Emukule G, Angala P, Mwega I, et al. Increasing access to HIV counseling and testing through mobile services in Kenya: strategies, utilization, and cost-effectiveness. *J Acquir Immune Defic Syndr* 2010 Jul;54(3):317-323 [FREE Full text] [doi: [10.1097/QAI.0b013e3181ced126](https://doi.org/10.1097/QAI.0b013e3181ced126)] [Medline: [20453819](https://pubmed.ncbi.nlm.nih.gov/20453819/)]
18. Nglazi MD, van Schaik N, Kranzer K, Lawn SD, Wood R, Bekker L. An incentivized HIV counseling and testing program targeting hard-to-reach unemployed men in Cape Town, South Africa. *J Acquir Immune Defic Syndr* 2012 Mar 1;59(3):e28-e34 [FREE Full text] [doi: [10.1097/QAI.0b013e31824445f0](https://doi.org/10.1097/QAI.0b013e31824445f0)] [Medline: [22173039](https://pubmed.ncbi.nlm.nih.gov/22173039/)]
19. Bassett IV, Regan S, Luthuli P, Mbonambi H, Bearnot B, Pendleton A, et al. Linkage to care following community-based mobile HIV testing compared with clinic-based testing in Umlazi Township, Durban, South Africa. *HIV Med* 2014 Jul;15(6):367-372 [FREE Full text] [doi: [10.1111/hiv.12115](https://doi.org/10.1111/hiv.12115)] [Medline: [24251725](https://pubmed.ncbi.nlm.nih.gov/24251725/)]
20. van Schaik N, Kranzer K, Wood R, Bekker LG. Earlier HIV diagnosis--are mobile services the answer? *S Afr Med J* 2010 Oct;100(10):671-674. [Medline: [21080998](https://pubmed.ncbi.nlm.nih.gov/21080998/)]
21. HIV-SAM Project. Vrijwillig hiv-testen en counselen (VCT) in Afrikaanse gemeenschapssettings. Project report 30 juni 2010. Antwerp: Institute of Tropical Medicine; 2010. URL: <http://www.itg.be/itg/Uploads/Klinische%20wetenschappen/HIVSAM/Rapport%20VCT%20in%20Afrikaanse%20gemeenschapssettings.pdf> [accessed 2016-01-11] [WebCite Cache ID 6eSVdhXX4]
22. Navaza B, Guionnet A, Navarro M, Estévez L, Pérez-Molina JA, López-Vélez R. Reluctance to do blood testing limits HIV diagnosis and appropriate health care of sub-Saharan African migrants living in Spain. *AIDS Behav* 2012 Jan;16(1):30-35. [doi: [10.1007/s10461-010-9871-0](https://doi.org/10.1007/s10461-010-9871-0)] [Medline: [21259044](https://pubmed.ncbi.nlm.nih.gov/21259044/)]
23. Fransen K, Vermoesen T, Beelaert G, Menten J, Hutse V, Wouters K, et al. Using conventional HIV tests on oral fluid. *J Virol Methods* 2013 Dec;194(1-2):46-51. [doi: [10.1016/j.jviromet.2013.08.004](https://doi.org/10.1016/j.jviromet.2013.08.004)] [Medline: [23969313](https://pubmed.ncbi.nlm.nih.gov/23969313/)]
24. Pai NP, Balram B, Shivkumar S, Martinez-Cajas JL, Claessens C, Lambert G, et al. Head-to-head comparison of accuracy of a rapid point-of-care HIV test with oral versus whole-blood specimens: a systematic review and meta-analysis. *The Lancet Infectious Diseases* 2012 May;12(5):373-380. [doi: [10.1016/S1473-3099\(11\)70368-1](https://doi.org/10.1016/S1473-3099(11)70368-1)]
25. Ellen JM, Liang TS, Jacob CA, Erbeling E, Christmyer C. Post-HIV test counselling of clients of a mobile STD/HIV clinic. *Int J STD AIDS* 2004 Nov;15(11):728-731. [doi: [10.1258/0956462042395258](https://doi.org/10.1258/0956462042395258)] [Medline: [15537457](https://pubmed.ncbi.nlm.nih.gov/15537457/)]
26. Castro D, Quatremère G, Le Gall JM. HIV-COBATEST Project steering committee. Implementation of community based Voluntary Counseling and Testing (VCT) programs and services. Qualitative study report. 2013. URL: [http://balthiv.com/files\\_and\\_documents/Doc2.pdf](http://balthiv.com/files_and_documents/Doc2.pdf) [accessed 2016-01-11] [WebCite Cache ID 6eSW11LVB]
27. Platteau T, Fransen K, Apers L, Kenyon C, Albers L, Vermoesen T, et al. Swab2know: An HIV-Testing Strategy Using Oral Fluid Samples and Online Communication of Test Results for Men Who Have Sex With Men in Belgium. *J Med Internet Res* 2015;17(9):e213 [FREE Full text] [doi: [10.2196/jmir.4384](https://doi.org/10.2196/jmir.4384)] [Medline: [26330138](https://pubmed.ncbi.nlm.nih.gov/26330138/)]
28. Fransen K, Vermoesen T, Beelaert G, Menten J, Hutse V, Wouters K, et al. Using conventional HIV tests on oral fluid. *Journal of Virological Methods* 2013 Dec;194(1-2):46-51. [doi: [10.1016/2013.08.004](https://doi.org/10.1016/2013.08.004)]
29. DeWalt K, De Walt BR. Participant Observation: A Guide for Fieldworkers. Walnut Creek: AltaMira Press; 2002.
30. May T. Social Research: Issues, Methods and Research. Glasgow: Open University Press; 2011.
31. Patton MQ. Qualitative research and evaluation methods. 3rd edition. Thousand Oaks, California: Sage Publications; 2002.
32. McCoy HV, Hlaing W, Ergon-Rowe E, Samuels D, Malow R. Lessons from the fields: a migrant HIV prevention project. *Public Health Rep* 2009;124(6):790-796 [FREE Full text] [Medline: [19894420](https://pubmed.ncbi.nlm.nih.gov/19894420/)]
33. Williams TT, Pichon LC, Latkin CA, Davey-Rothwell M. Practicing What is Preached: The Relationship between Congregational Support and HIV Prevention Behaviors among Black Women. *J Community Psychol* 2014 Apr;42(3):365-378 [FREE Full text] [doi: [10.1002/jcop.21615](https://doi.org/10.1002/jcop.21615)] [Medline: [25435597](https://pubmed.ncbi.nlm.nih.gov/25435597/)]
34. Fakoya I, Reynolds R, Caswell G, Shiripinda I. Barriers to HIV testing for migrant black Africans in Western Europe. *HIV Med* 2008 Jul;9 Suppl 2:23-25 [FREE Full text] [doi: [10.1111/j.1468-1293.2008.00587.x](https://doi.org/10.1111/j.1468-1293.2008.00587.x)] [Medline: [18557866](https://pubmed.ncbi.nlm.nih.gov/18557866/)]
35. Gama A, Fraga S, Dias S. Impact of socio-demographic factors on HIV testing among African immigrants in Portugal. *J Immigr Minor Health* 2010 Dec;12(6):841-846. [doi: [10.1007/s10903-010-9324-x](https://doi.org/10.1007/s10903-010-9324-x)] [Medline: [20157850](https://pubmed.ncbi.nlm.nih.gov/20157850/)]
36. Sadler KE, McGarrigle CA, Elam G, Ssanyu-Sseruma W, Davidson O, Nichols T, et al. Sexual behaviour and HIV infection in black-Africans in England: results from the Mayisha II survey of sexual attitudes and lifestyles. *Sex Transm Infect* 2007 Dec;83(7):523-529 [FREE Full text] [doi: [10.1136/sti.2007.027128](https://doi.org/10.1136/sti.2007.027128)] [Medline: [17932129](https://pubmed.ncbi.nlm.nih.gov/17932129/)]
37. UNAIDS, GNP+. Positive Health, Dignity and Prevention: Operational Guidelines. Geneva and Amsterdam; 2013. URL: [http://www.unaids.org/sites/default/files/media\\_asset/20130802\\_Positive\\_Health\\_Dignity\\_Prevention\\_Operational\\_Guidelines\\_0.pdf](http://www.unaids.org/sites/default/files/media_asset/20130802_Positive_Health_Dignity_Prevention_Operational_Guidelines_0.pdf) [WebCite Cache ID 6eTgJJrld]



38. Ndiaye B, Salleron J, Vincent A, Bataille P, Bonnevie F, Choisy P, et al. Factors associated with presentation to care with advanced HIV disease in Brussels and Northern France: 1997-2007. *BMC Infect Dis* 2011;11:11 [[FREE Full text](#)] [doi: [10.1186/1471-2334-11-11](https://doi.org/10.1186/1471-2334-11-11)] [Medline: [21226905](#)]
39. Prost A, Sseruma WS, Fakoya I, Arthur G, Taegtmeier M, Njeri A, et al. HIV voluntary counselling and testing for African communities in London: learning from experiences in Kenya. *Sex Transm Infect* 2007 Dec;83(7):547-551 [[FREE Full text](#)] [doi: [10.1136/sti.2007.027110](https://doi.org/10.1136/sti.2007.027110)] [Medline: [17911136](#)]
40. Federal Public Service Health, Food Chain Security and Environment. National HIV Plan Belgium 2014-2019. 2013. URL: <http://www.breach-hiv.be/media/docs/HIVPlan/NationalPlanDutch.pdf> [[WebCite Cache ID 6eSX8DBFB](#)]

## Abbreviations

**GP:** general practitioner  
**HIV:** human immunodeficiency virus  
**LGBT:** lesbian Gay Bisexual Transsexuals  
**MSM:** men who have sex with men  
**STI:** Sexually Transmitted Infections  
**VCT:** voluntary HIV counseling and testing

*Edited by G Eysenbach; submitted 12.01.16; peer-reviewed by A McNaghten, K Delaney; comments to author 18.04.16; revised version received 11.05.16; accepted 17.06.16; published 04.08.16.*

*Please cite as:*

Loos J, Manirankunda L, Platteau T, Albers L, Fransen K, Vermoesen T, Namanya F, Nöstlinger C  
*Acceptability of a Community-Based Outreach HIV-Testing Intervention Using Oral Fluid Collection Devices and Web-Based HIV Test Result Collection Among Sub-Saharan African Migrants: A Mixed-Method Study*  
*JMIR Public Health Surveill* 2016;2(2):e33  
URL: <http://publichealth.jmir.org/2016/2/e33/>  
doi: [10.2196/publichealth.5519](https://doi.org/10.2196/publichealth.5519)  
PMID: [27493067](https://pubmed.ncbi.nlm.nih.gov/27493067/)

©Jasna Loos, Lazare Manirankunda, Tom Platteau, Laura Albers, Katrien Fransen, Tine Vermoesen, Fiona Namanya, Christiana Nöstlinger. Originally published in *JMIR Public Health and Surveillance* (<http://publichealth.jmir.org>), 04.08.2016. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in *JMIR Public Health and Surveillance*, is properly cited. The complete bibliographic information, a link to the original publication on <http://publichealth.jmir.org>, as well as this copyright and license information must be included.

Original Paper

# Determining Survey Satisficing of Online Longitudinal Survey Data in the Multicenter AIDS Cohort Study: A Group-Based Trajectory Analysis

Junrui Di<sup>1</sup>, MSc; Ying Li<sup>1</sup>, MSc, PhD; M Reuel Friedman<sup>2</sup>, PhD; Susheel Reddy<sup>3</sup>, MPH; Pamela J Surkan<sup>4</sup>, PhD, ScD; Steven Shoptaw<sup>5</sup>, PhD; Michael Plankey<sup>1</sup>, PhD

<sup>1</sup>Georgetown University Medical Center, Department of Medicine, Division of Infectious Diseases, Washington, DC, United States

<sup>2</sup>University of Pittsburgh, School of Public Health, Department of Infectious Diseases and Microbiology, Pittsburgh, PA, United States

<sup>3</sup>Northwestern University, Feinberg School of Medicine, Division of Infectious Disease, Chicago, IL, United States

<sup>4</sup>Johns Hopkins University, Bloomberg School of Public Health, Social and Behavioral Interventions Program, Baltimore, MD, United States

<sup>5</sup>University of California-Los Angeles, David Geffen School of Medicine, Department of Family Medicine, Los Angeles, CA, United States

**Corresponding Author:**

Michael Plankey, PhD

Georgetown University Medical Center

Department of Medicine, Division of Infectious Diseases

2115 Wisconsin Avenue, NW, Suite 130

Washington, DC, 20007

United States

Phone: 1 202784260

Fax: 1 2027840791

Email: [mwp23@georgetown.edu](mailto:mwp23@georgetown.edu)

## Abstract

**Background:** Survey satisficing occurs when participants respond to survey questions rapidly without carefully reading or comprehending them. Studies have demonstrated the occurrence of survey satisficing, which can degrade survey quality, particularly in longitudinal studies.

**Objective:** The aim of this study is to use a group-based trajectory analysis method to identify satisficers when similar survey questions were asked periodically in a long-standing cohort, and to examine factors associated with satisficing in the surveys having sensitive human immunodeficiency virus (HIV)-related behavioral questions.

**Methods:** Behavioral data were collected semiannually online at all four sites of the Multicenter AIDS Cohort Study (MACS) from October 2008 through March 2013. Based on the start and end times, and the word counts per variable, response speed (word counts per second) for each participant visit was calculated. Two-step group-based trajectory analyses of the response speed across 9 study visits were performed to identify potential survey satisficing. Generalized linear models with repeated measures were used to investigate the factors associated with satisficing on HIV-related behavioral surveys.

**Results:** Among the total 2138 male participants, the median baseline age was 51 years (interquartile range, 45-58); most of the participants were non-Hispanic white (62.72%, 1341/2138) and college graduates (46.59%, 996/2138), and half were HIV seropositive (50.00%, 1069/2138). A total of 543 men (25.40%, 543/2138) were considered potential satisficers with respect to their increased trajectory tendency of response speed. In the multivariate analysis, being 10 years older at the baseline visit increased the odds of satisficing by 44% (OR 1.44, 95% CI 1.27-1.62,  $P<.001$ ). Compared with the non-Hispanic white participants, non-Hispanic black participants were 122% more likely to satisfice the HIV-related behavioral survey (OR 2.22, 95% CI 1.69-2.91,  $P<.001$ ), and 99% more likely to do so for the other race/ethnicity group (OR 1.99, 95% CI 1.39-2.83,  $P<.001$ ). Participants with a high school degree or less were 67% more likely to satisfice the survey (OR 1.67, 95% CI 1.26-2.21,  $P<.001$ ) compared with those with a college degree. Having more than one sex partner and using more than one recreational drug reduced the odds of satisficing by 24% (OR 0.76, 95% CI 0.61-0.94,  $P=.013$ ) and 28% (OR 0.72, 95% CI 0.55-0.93,  $P=.013$ ), respectively. No statistically significant association of HIV serostatus with satisficing was observed.

**Conclusions:** Using a group-based trajectory analysis method, we could identify consistent satisficing on HIV-related behavioral surveys among participants in the MACS, which was associated with being older, being non-white, and having a lower education

level; however, there was no significant difference by HIV serostatus. Methods to minimize satisficing using longitudinal survey data are warranted.

(*JMIR Public Health Surveill* 2016;2(2):e150) doi:[10.2196/publichealth.5240](https://doi.org/10.2196/publichealth.5240)

## KEYWORDS

ACASI; cohort studies; data collection; data quality; group-based trajectory analysis; reading speed; survey satisficing

## Introduction

Surveys have been widely used to collect data and provide important information for a wide variety of fields, including medical research [1]. However, the growing popularity of surveys to answer questions has led to a tendency to overlook the fact that surveys can involve many technical problems that can affect the accuracy and consistency of the data collected [2].

Satisficing is a particularly important and complicated issue for survey design that involves various psychological factors and attitude measurements. In an optimized survey administration, respondents will fully engage in four steps (1) interpreting the question; (2) searching memory for relevant information; (3) integrating information into summary judgment; and (4) reporting judgment [3-8]. The opposite of survey optimizing is referred to as “survey satisficing”, whereby a person engages in steps 2 or 3 half-heartedly or even skips steps 2 or 3 [3]. To be more specific, survey satisficing occurs when participants complete the questions rapidly without reading or comprehending them [8], or as described by Krosnick [9], it is the tendency of respondents to provide satisfactory but not optimal answers in order to reduce their effort. Garland et al [10] have shown with an interesting example that the survey responses containing both true answers and satisficed answers demonstrated divergent results. Therefore, learning about the presence and patterns of satisficing will help us design better surveys to maximize the quality of the responses received [11].

In 1991, Krosnick [9] built up a fundamental theory of satisficing, which has been widely referred. In his theory, he categorized satisficing into weak and strong satisficing, and he listed response strategies that were highly likely to cause satisficing and conditions that would foster satisficing. Depending on the format and features of the survey questions and response options, Vannette and Krosnick [12] categorized reasons for respondents’ satisficing into four responses, saying “don’t know” when they do know or could know, “acquiescence” (ie, agreeing with statements with which they might actually disagree), “response order effects”, (ie, respondents may be inclined to settle for the first plausible response option they identify when they are offered nominal or ordinal response choices), and “nondifferentiation in using rating scales” (ie, so-called straightlining when more thought might lead to different answers for different statements) [12]. In addition, Vannette provided a solution to design a survey to make it more difficult for respondents to engage in satisficing [13]. Four satisficing metrics have been used by Barge and Gehlbach [14] to demonstrate the measurable effect on survey results, which included “early termination”, “nondifferentiation”, “skipping items”, and “rushing”. An internal study performed

by the technical staff at SurveyMonkey showed that a higher number of questions on a survey was highly correlated with faster speed answering each question, decreasing the validity of the data [15]. Kapelner and Chandler [16] suggested two treatments to prevent satisficing in online surveys: “timing control” (ie, featuring a disabled continue button for the duration of the waiting period) and “Kapcha” (ie, having a waiting period and attempting to attract the attention of respondents by sequentially “fading in” each word in the question’s directions, its prompt, and its answer choices). Daniel [8] found that questions toward the end of long surveys were more likely to attract a satisficing response, and one strategy to minimize this effect was to randomize question order without considering other effects of question order. Holbrook et al [17] linked response heaping (also referred to as “rounding” or “digit preference”) to satisficing.

Satisficing identification methods have been developed based on examining the occurrence of such “suspicious” response strategies highly associated with satisficing that Krosnick [9] listed, such as “don’t know”, “acquiescence”, and “straightlining” [12] under the assumption that more such behaviors indicate more satisficing and lower response quality. Oppenheimer et al [18] developed a methodological tool, termed “instructional manipulation check”, which provided an indirect measure of satisficing by measuring whether or not participants read the instructions. Garland et al [10] proposed a Bayesian inference approach that examined survey responses in quantitative form to recognize a “normal” pattern of results. Turner et al [19] recommended using response latencies to assess whether the respondents minimized their cognitive costs.

However, all of these reported satisficing theories, findings, and indicators used for satisficing identification focus on a single survey. For a longitudinal study in which surveys with similar sensitive questions are performed periodically, the reasons for respondents’ satisficing are more complicated. It is highly possible that participants in a longitudinal study become more familiar with the structure and response pattern of the survey questionnaire after several visits, and satisficing may occur due to this learning pattern compounded by participants’ lack of patience with carefully reading questions. Such a problem has been reported in a previous study [20], where many respondents complained that they were asked repeatedly to report detailed information that did not frequently change. Therefore, new measurable parameter(s) used to identify satisficing on multiple-visit surveys need to be investigated in a longitudinal study.

Literature has shown that under limited time conditions, as reading speed increases, the comprehension and recognition of important and unimportant information from a text subsequently deteriorates [21-24]. Moreover, for a survey questionnaire, it is

more important to study the response speeding (ie, giving answers really quickly). Response speeding has been proven to be associated with “suspicious” response strategies, so that can be considered a parameter for satisficing identification. As pointed out by Krosnick [9], the longer an interview has been under way, the lower the respondent’s motivation to optimize and the more likely satisficing were to flourish. The “rushing” metrics Barge and Gehlbach [14] mentioned were essentially capturing faster response speed. Zhang and Conrad [25] suggested speeding was positively related to straightlining. Malhotra [26] found that speeding made respondents more likely to choose the options presented earlier regardless of the content. Conrad et al [27] observed an association between speeding and straightlining across a relatively small set of items. Wells et al [28] found that respondents who engage more in speeding were less likely to choose “other” and to elaborate on their answers. Bathelt and Bauknecht [29] compared computer-assisted personal interviewing (CAPI) and computer-assisted telephone interviewing (CATI) interviews with respect to the effect of speeding on satisficing and discovered that speeding was positively associated with satisficing in CAPI.

As pointed out by Zhang and Conrad [25], speeding in surveys is not only associated with “faster” responses, but the responses are also given faster than an appropriate speed threshold that was set low enough to capture answers that are unreasonably fast. However, this threshold is usually calculated based on a single survey questionnaire [14,25], which ignores the variation of survey questionnaires across visit times in a survey-based longitudinal study. Thus, using a universal threshold to identify speeding might not be appropriate in a longitudinal study.

During a Multicenter AIDS Cohort Study (MACS) semiannual visit, study participants are asked to complete an audio computer-assisted self-interviewing (ACASI) behavioral survey. Based on the study by Tourangeau and Yan [30], survey questions about illicit drug use and sexual behaviors are considered sensitive and tend to produce poor data quality. As an ongoing longitudinal study of 30 years targeting the cohort of homosexual men with human immunodeficiency virus (HIV) infection and with major involvement of sensitive questions about illicit drug use, sexual behaviors, and income, it is unclear whether or not satisficing has occurred with the MACS ACASI survey data over visit times, which can decrease response quality. Therefore, this current study aims to (1) plot the grouped trajectory of the MACS participants’ response speeds with visit times; (2) identify the MACS ACASI survey satisficers based on the trajectory analysis results of the longitudinal data; and (3) investigate the factors associated with satisficing in the MACS surveys having HIV-related highly sensitive questions.

## Methods

### Recruitment

The MACS is an ongoing prospective study of the natural and treated histories of HIV infection among men who have sex with men in the United States. A total of 6972 men were recruited (4954 in 1984-1985, 668 in 1987-1991, and 1350 in 2001-2003) at four centers located in Baltimore, MD/Washington, DC, Chicago, IL, Los Angeles, CA, and

Pittsburgh, PA. The study design of the MACS has been described in detail previously [31,32], and only methods relevant to the current study are presented here. All MACS questionnaires are available on the MACS website [33]. The MACS study protocols were approved by the institutional review boards of each of the participating sites, their community partners, and community advisory boards, and informed consent was obtained from all participants. MACS participants return every 6 months for detailed interviews (ie, an ACASI survey), physical examinations, and collection of blood for laboratory testing and storage in a central repository.

In this study, we collected behavioral survey data via ACASI at all four MACS sites across 9 visits (ie, MACS visits 50-58) from October 1, 2008 to March 31, 2013. The survey had 7 sections including approximately 200 questions (since visit 56 from October 1, 2011, a short questionnaire with approximately 30 fewer questions was used) related to participant’s socio-demographics, illicit drug use, alcohol and cigarettes use, sexual activity, quality of life, pre-exposure or post-exposure prophylaxis, and sexual health since the previous visit.

### Measures

For each participant visit, the total survey time was calculated as the difference between the start and end times collected automatically by the computer system during the ACASI survey, and the total number of words of responded variables were counted. Some questions could be skipped logically depending on the responses to the previous question(s). For example, when a respondent answered “no” to the question “have you engaged in any sexual activities involving another person since your last visit”, the following 50 or more questions about sexual activity would be skipped. Unskipped questions could not be ignored; that is, participants had to give a response to an unskipped question before proceeding to the next one.

The primary outcome of interest was the response speed, which was calculated with the following formula:

$$\text{Response Speed} = N/T$$

where T is the total survey time (in seconds) and N is the total number of word counts of a responded variable. Note that the speed was calculated based on word count of a “variable” instead of word counts of a “question” because some survey question(s) corresponded to more than one variable.

Socio-demographics including race/ethnicity (non-Hispanic white, non-Hispanic black, or others), age at baseline visit (ie, visit 50), education level (less than a college degree, college degree, or higher than a college degree), HIV serostatus (HIV+ or HIV-), sexual activity (ie, had no sex since last visit, had one sexual partner since last visit, or had more than one sexual partner since last visit), and recreational drug use (ie, not used any illicit drug since last visit, used only one kind of illicit drug since last visit, used more than one kind of illicit drug since last visit) were studied to explore the factors associated with satisficing.

### Statistical Analysis

Group-based trajectory analyses [34] modeling the participant’s response speed across the 9 study visits were performed to



define participants' survey satisficing status. The model selection for the trajectory analysis was based on the largest negative Bayesian information criteria [35]. Based on the trajectory analysis results, a participant's survey satisficing status (ie, satisficing or nonsatisficing) was defined. That is, the participants grouped with an increasing tendency of the response speed across the 9 visits via the trajectory analyses were considered satisficers. Univariate and multivariate generalized linear models with repeated measures were run to examine the associations of socio-demographic characteristics with survey satisficing. The iterative fitting algorithm was used for repeated measures in modeling to avoid the violation of the assumption of independence due to the multiple visits of the same participants. Statistical significance was evaluated at the .05 level, and ORs and 95% CIs were calculated. All analyses were performed using SAS version 9.3 (SAS Institute, Inc., Cary, NC, USA).

## Results

### Identification of Satisficers

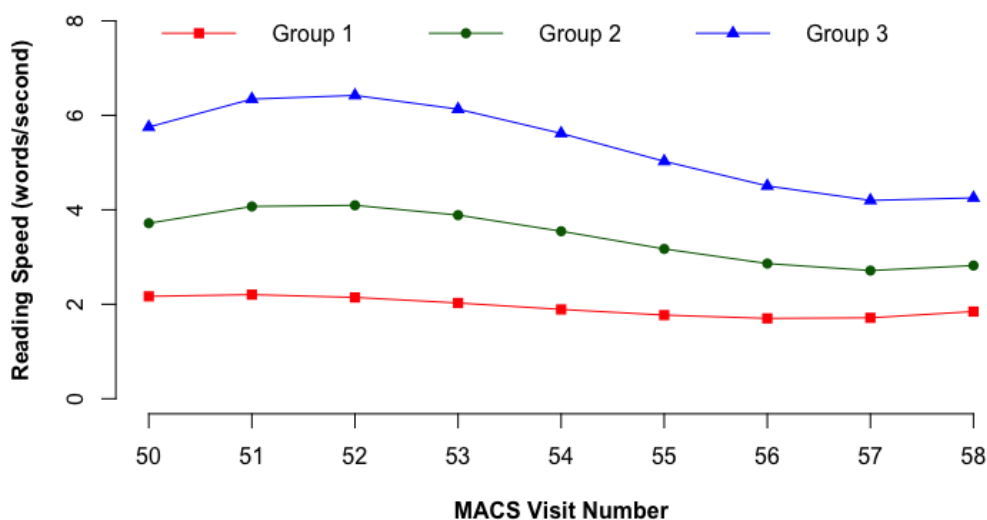
There were a total of 14,722 participant visits from visits 50 to 58. After the exclusion of 14 participant visits without a start time, 1887 visits missing end times, and another 34 visits with illogical times recorded (eg, start time was 9:00 pm and the end time was 10:00 am), the study sample was composed of 12,787 participant visits contributed by 2138 MACS participants.

The results of two-step trajectory analyses of the participants' response speed across 9 MACS visits are plotted in Figures 1

and 2. The group-based analysis among the total of 2138 MACS participants yielded a solution that identified three groups (see Figure 1): (1) a group whose speed was approximately 6 words per second and had a decreased tendency with time (group 3; 13.14% of the cohort, 281/2138); (2) a group with a speed of approximately 4 words per second (group 2; 49.01% of the cohort, 1048/2138) showing a decreased and then constant pattern, and (3) a group whose speed of approximately 2 words per second seemed constant for the first 7 visits and then slightly increased across the recent 3 visits (group 1; 37.84% of the cohort, 809/2138). Although for group 3 response speed appeared to increase from visit 50 to visit 52, we did not consider them to be satisficing because the trajectory showed an overall decreasing pattern.

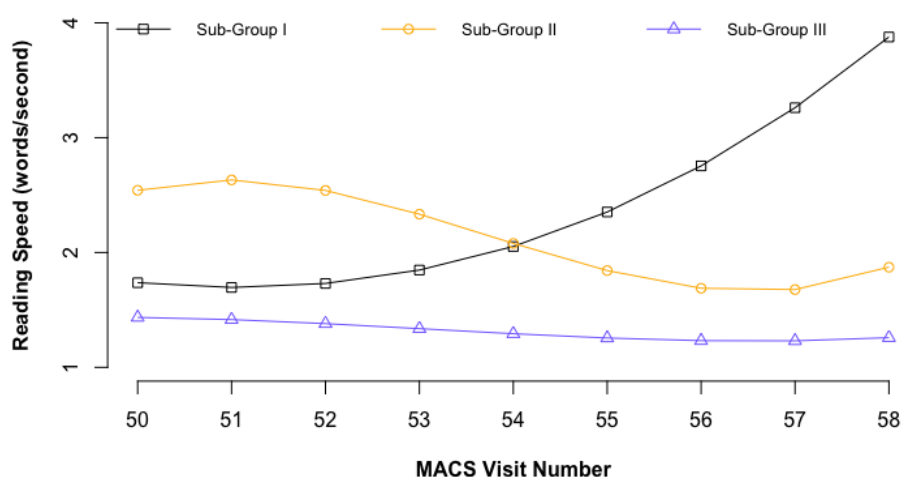
Because only group 1 showed a pattern with increased response speed during recent visits, the survey satisficers most likely existed in group 1. To observe more clear patterns, further trajectory analyses were conducted in group 1. The results of second-step trajectory analyses are plotted in Figure 2, in which three subgroups were identified. Subgroup III (32.88% of group 1, 266/809) had a nearly constant speed; subgroup I (4.33% of group 1, 35/809) revealed a dramatically increasing trajectory in response speed; and subgroup II (62.79% of group 1, 508/809) showed a decreasing pattern up to visit 56 and an accelerated speed after visit 56. Due to the accelerated response speed across the recent visits, participants in subgroups I and II were both identified as satisficers. Therefore, among the 2138 participants, 543 (25.40%, 543/2138) were identified as satisficers, and the remaining 1595 (74.60%, 1595/2138) were nonsatisficers.

**Figure 1.** First-step trajectory analysis of response speed (word counts per second) across 9 MACS visits among the total 2138 MACS participants.





**Figure 2.** Second-step trajectory analysis of response speed (word counts per second) across 9 MACS visits among the 809 participants in group 1 of Figure 1.



### Characteristics of the Study Population by Satisficing Status

Among the total 2138 participants, the median age at baseline visit was 51 years; most were non-Hispanic white (62.72%, 1341/2138) and college graduates (46.59%, 996/2138), half were HIV seropositive (50.00%, 1069/2138), and 543 (25.40%, 543/2138) were identified as survey satisficers (Table 1). Among

the 543 satisficers, 49.36% (268/543) were non-Hispanic white, 35.73% (194/543) non-Hispanic black, 54.70% (297/543) HIV seropositive, and 73.30% (398/543) had a college degree or higher. Among the 1595 nonsatisficers, 67.27% (1073/1595) were non-Hispanic white, 21.25% (339/1595) non-Hispanic black, 48.40% (772/1595) HIV seropositive, and 85.08% (1357/1595) had a college degree or higher.

**Table 1.** Characteristics of the study population by satisficing status.

Characteristics	Satisficing status, n (%)		Total, n (%)
	Nonsatisficer	Satisficer	
N	1595 (74.60)	543 (25.40)	2138 (100.00)
Baseline age in years, median (IQR <sup>a</sup> )	51 (45, 57)	52 (45, 59)	51 (45, 58)
Race			
Non-Hispanic white	1073 (67.27)	268 (49.36)	1341 (62.72)
Non-Hispanic black	339 (21.25)	194 (35.73)	533 (24.93)
Other	183 (11.47)	81 (14.92)	264 (12.35)
Education			
Less than college	238 (14.92)	145 (26.70)	383 (17.91)
College	756 (47.40)	240 (44.20)	996 (46.59)
Higher than college	601 (37.68)	158 (29.10)	759 (35.50)
HIV <sup>b</sup> serostatus			
Negative	823 (51.60)	246 (45.30)	1069 (50.00)
Positive	772 (48.40)	297 (54.70)	1069 (50.00)

<sup>a</sup>IQR: interquartile range.

<sup>b</sup>HIV: human immunodeficiency virus.

### Factors Associated With Satisficing

In the univariate analysis, age at baseline visit, race/ethnicity, education, sexual activity, and drug use were statistically significantly associated with satisficing (Table 2). In particular, with a 10-year increase of baseline age, the odds of satisficing increased by 16% (OR 1.16, 95% CI 1.05-1.29,  $P < .001$ ). Non-Hispanic black men and those of other race/ethnicity groups

were 117% (OR 2.17, 95% CI 1.71-2.76,  $P < .001$ ) and 60% (OR 1.60, 95% CI 1.16-2.20,  $P < .001$ ) more likely to satisfice the survey, respectively, compared with non-Hispanic white participants. Compared with those with a college degree, the odds of satisficing were 29% lower (OR 0.71, 95% CI 0.61-0.83,  $P < .001$ ) for participants with a high school degree or less and 15% lower (OR 0.85, 95% CI 0.74-0.98,  $P = .02$ ) for those with a graduate degree. Compared with those without sexual activity,

having one sex partner decreased the odds of satisficing by 26% (OR 0.74, 95% CI 0.61-0.91,  $P<.001$ ), and having more than one sex partner decreased the odds by 40% (OR 0.60, 95% CI 0.49-0.73,  $P<.001$ ). Last, the odds of satisficing were 35% lower

for participants who used more than one recreational drug (OR 0.65, 95% CI 0.51-0.83,  $P<.001$ ) than those without drug use. No statistically significant association of HIV serostatus with satisficing was observed.

**Table 2.** Odds ratios with 95% confidence intervals and  $P$  values of satisficing estimated with univariate and multivariate models.

	Univariate model		Multivariate model	
	OR (95% CI)	$P$	OR (95% CI)	$P$
Socio-demographics				
Baseline age, 10-year increase	1.16 (1.05-1.29)	.004	1.44 (1.27-1.62)	<.001
Race/ethnicity				
Non-Hispanic white	1 (Reference)		1 (Reference)	
Non-Hispanic black	2.17 (1.71-2.76)	<.001	2.22 (1.69-2.91)	<.001
Other	1.60 (1.16-2.20)	.004	1.99 (1.39-2.83)	<.001
Education				
Less than a college degree	0.71 (0.61-0.83)	<.001	1.67 (1.26-2.21)	<.001
College degree	1 (Reference)		1 (Reference)	
Higher than a college degree	0.85 (0.74-0.98)	.02	0.83 (0.65-1.07)	.16
HIV <sup>a</sup> serostatus				
Negative	1 (Reference)		1 (Reference)	
Positive	1.23 (0.99-1.52)	.054	1.13 (0.90-1.42)	.29
Sexual activity				
No sexual activity	1 (Reference)		1 (Reference)	
Had one sex partner	0.74 (0.61-0.91)	.004	0.85 (0.69-1.05)	.14
Had more than one partner	0.60 (0.49-0.73)	<.001	0.76 (0.61-0.94)	.013
Recreational drug use				
No drug use	1 (Reference)		1 (Reference)	
Used only one drug	0.92 (0.76-1.12)	.42	1.02 (0.84-1.25)	.84
Used more than one drug	0.65 (0.51-0.83)	<.001	0.72 (0.55-0.93)	.013

<sup>a</sup>HIV: human immunodeficiency virus.

After considering all socio-demographic factors in a model (Table 2), a 10-year increase of baseline age increased the odds of satisficing by 44% (OR 1.44, 95% CI 1.27-1.62;  $P<.001$ ). Compared with non-Hispanic white participants, non-Hispanic black men were 122% times more likely to satisfice the survey (OR 2.22, 95% CI 1.69-2.91,  $P<.001$ ), and the other race/ethnicity group was 99% more likely to satisfice (OR 1.99, 95% CI 1.39-2.83,  $P<.001$ ). Participants with a high school degree or less were 67% more likely (OR 1.67, 95% CI 1.26-2.21,  $P<.001$ ) to satisfice the survey in comparison with those with a college degree. Having more than one sex partner and using more than one drug reduced the odds of satisficing by 24% (OR 0.76, 95% CI 0.61-0.94,  $P=.013$ ) and 28% (OR 0.72, 95% CI 0.55-0.93,  $P=.013$ ), respectively. We did not observe any statistically significant difference in satisficing by HIV status.

## Discussion

### Principal Findings

We studied a method using group-based trajectory analysis to investigate satisficing on the multiple-visit surveys in a longitudinal study, which would be supplementary to the traditional Krosnick's theory focusing on a single survey. We found 25.40% (543/2138) of MACS participants were survey satisficing over time. Factors such as older age, less education, and non-Hispanic black or other race/ethnicity were positively associated with satisficing on the surveys having HIV-related sensitive questions, whereas being sexually active and using more than one recreational drug were negatively associated with satisficing. No statistically significant differences in satisficing by HIV serostatus were found.

### Comparison With Prior Work

In the fundamental Krosnick [9] theory, it was found that when question alternatives were presented to respondents visually, weak satisficing was likely to produce primacy effects. MACS used the ACASI survey, where question alternatives were

visually presented to the respondent, which was potentially one source of satisficing. Moreover, in the MACS ACASI survey, when respondents were asked about quality of life and illicit drug use, the questionnaire was designed to rate a series of objects on a common scale. For example, when respondents were asked about how they feel and how things have been going during the past 4 weeks, the questionnaires just provided a list of such questions with the same scale (eg, “all of the time”, or “most of the time”). Discussed by Krosnick’s theory, this kind of questionnaire design could cause respondents to fail to differentiate between items and give all or almost all of the objects the same rating (ie, “straightlining”), and it might lead to strong satisficing [9,36].

From a survey-quality perspective, speeding has been shown to be associated with behaviors always considered indicators of survey satisficing and poor response quality [25-29]. On one hand, speeders are likely to be professional respondents solely motivated by incentives offered for survey completion and unconcerned with the survey itself or with the answers they provide. On the other hand, speeders may also be well-intentioned respondents who get frustrated with a survey (eg, too long, boring topic, contains lengthy grid items, or requires answers for every item) and react by speeding through the survey. In either case, speeding is considered problematic survey behavior because respondents are not providing thoughtful, accurate answers. One noticeable difference between speeders and non-speeders was with responses to open-ended times (questions with text boxes provided for qualitative responses), and speeders are more likely to skip these questions and not provide a response [37]. Zhang and Conrad [25] used a predefined speed threshold to detect speeding. Unfortunately, all current theories and methods about satisficing are based on a single survey, which would be not appropriate for the complicated case of a longitudinal study with several surveys on multiple visits. In this study, we used group-based trajectory analysis, a statistical approach based on mixture models that can model unobserved heterogeneity in population, to study the pattern of the trajectory of participants’ average response speed across 9 visits in a longitudinal study and to identify survey satisficing through the objective tendency of participants’ speeding across time. Moreover, our study provided abundant information regarding satisficing on such a highly sensitive survey questionnaire containing HIV-related questions for illicit drug use and sexual behaviors among men who have sex with men, those who are at high risk for HIV infection, and those who usually have poor response quality.

Data from the SurveyMonkey study demonstrated that more questions are highly associated with faster speed [15], which was in accordance with our observations at MACS after visit 56 with fewer survey questions. Less education (ie, high school degree or less) was associated with more satisficing, echoing previous studies [25-27,29,38]; as Krosnick [9] pointed out, the

possible reason is that lower-status respondents were likely to agree with any assertions that the interviewer or researcher apparently believes. We found that older people were more likely to satisfice. We think it was because the elderly can become fatigued sooner, which is supported by Bathelt’s findings using the CAPI interview [29], yet differed from results of the studies by Zhang and Conrad [25] and Beckers et al [39]. Krosnick [9] stated that respondent motivation to optimize a survey (which is also a factor associated with satisficing) was influenced by the degree to which the topic of a question was personally important to the respondent. Therefore, it makes sense that participants who reported sexual activity and recreational drug use were less likely to satisfice because this population in the MACS cohort usually has higher risk for HIV infection and they may pay more attention to the survey questionnaire to better protect themselves.

### Limitations

Certain limitations of this study deserve attention. There were approximately 10% of the records missing the end time and most of these records were from one study site. Self-administration of the questionnaire did not provide an opportunity for participants to ask questions about the survey, so it was not possible to verify the accuracy of the participants’ interpretation of the survey questions, which might affect their responses. In addition, the survey did not have an independent indication of the response speed of each participant as a benchmark before they completed the questionnaire. Except for the computer system automatically recording the start and end times of an online survey, there was no other time verification method in this study.

### Conclusions

This study provides valuable insight into the identification of survey satisficing in a long-standing observational cohort study where the patterns of repetitive question asking may have been learned over time, leading to satisficing, and factors associated with satisficing on surveys having HIV-related sensitive questions. Therefore, results from the longitudinal analyses of these behavioral data could be less precise over time, even after taking into account the within-participant variability over time. Technical solutions are needed to present survey questions so that respondents read and answer questions more carefully. Survey designers should consider some basic solutions to maximize respondent motivation (eg, keep the questionnaires short and place important questions early and include random probes such as “why do you say that?”), minimize task difficulty (eg, maximize the familiarity of the words and decompose questions whenever possible), minimize response effects (eg, avoid agree/disagree, true/false, and yes/no questions, avoid blocks of ratings on the same scale, and randomize the order of list of options) [11,13,40], and/or force the participant to see the question for a certain waiting period [16].

### Acknowledgments

Data in this manuscript were collected by the Multicenter AIDS Cohort Study (MACS). MACS (Principal Investigators): Johns Hopkins University Bloomberg School of Public Health (Joseph Margolick), U01-AI35042; Northwestern University (Steven

Wolinsky), U01-AI35039; University of California, Los Angeles (Roger Detels), U01-AI35040; University of Pittsburgh (Charles Rinaldo), U01-AI35041; and the Center for Analysis and Management of MACS, Johns Hopkins University Bloomberg School of Public Health (Lisa Jacobson), UM1-AI35043. The MACS is funded primarily by the National Institute of Allergy and Infectious Diseases, with additional co-funding from the National Cancer Institute, the National Institute on Drug Abuse, and the National Institute of Mental Health. Targeted supplemental funding for specific projects was also provided by the National Heart, Lung, and Blood Institute, and the National Institute on Deafness and Communication Disorders. MACS data collection is also supported by UL1-TR001079 (JHU ICTR) from the National Center for Advancing Translational Sciences, a component of the National Institutes of Health (NIH), and NIH Roadmap for Medical Research. The contents of this publication are solely the responsibility of the authors and do not represent the official views of the NIH, Johns Hopkins ICTR, or the National Center for Advancing Translational Sciences.

## Conflicts of Interest

None declared.

## References

1. Scheuren F. What is a Survey?. 2004. URL: <https://www.whatisasurvey.info/download.htm> [accessed 2014-08-27] [WebCite Cache ID 6S9DD7Lba]
2. Ferber R, Sheatsley P, Turner A, Waksberg J. What is a survey?. 1980. URL: <https://www.whatisasurvey.info/archive.htm> [accessed 2014-08-27] [WebCite Cache ID 6S9DFQ69Y]
3. Phil G. Satisficing: The Subtle Cancer of Surveys. 2010. URL: [https://www.surveymonkey.com/blog/en/blog/2010/10/04/satisficing\\_surveys/](https://www.surveymonkey.com/blog/en/blog/2010/10/04/satisficing_surveys/) [WebCite Cache ID 6S9DHdQLx]
4. Jabine TB, Straf ML, Tanur JM, Tourangeau R, editors. Report of the advanced research seminar on cognitive aspects of survey methodology. In: Cognitive Aspects of Survey Methodology: Building a Bridge Between Disciplines. Washington, DC: National Academy Press; 1984.
5. Tourangeau R, Rasinski KA. Cognitive processes underlying context effects in attitude measurement. Psychol Bull 1988;103(3):299-314. [doi: [10.1037/0033-2909.103.3.299](https://doi.org/10.1037/0033-2909.103.3.299)]
6. Tourangeau R. Attitude measurement: a cognitive perspective. In: Hippler HJ, Schwarz N, Sudman S, editors. Social Information Processing and Survey Methodology. New York, NY: Springer; 1987.
7. Cannell C, Miller P, Oksenberg L. Research on interviewing techniques. Sociol Methodol 1981;12(4):389-437. [doi: [10.2307/270748](https://doi.org/10.2307/270748)]
8. Daniel S. Satisficing in survey design. 2012 Presented at: Contemporary Approaches to Research in Mathematics, Science, Health and Environmental Education Symposium; 29-30 November 2012; Melbourne, Australia.
9. Krosnick JA. Response strategies for coping with the cognitive demands of attitude measures in surveys. Appl Cognit Psychol 1991 May;5(3):213-236. [doi: [10.1002/acp.2350050305](https://doi.org/10.1002/acp.2350050305)]
10. Garland P, Chen K, Epstein J, Suh A. Speed (necessarily) doesn't kill: a new way to detect survey satisficing. CASRO 2012;2012-2013:21 [FREE Full text]
11. Krosnick J. Survey Methods Newsletter. 2000. The threat of satisficing in surveys: the shortcuts respondents take in answering questions URL: <https://pprg.stanford.edu/wp-content/uploads/2000-The-threat-of-satisficing-in-surveys-The-shortcuts-responders.pdf> [accessed 2015-07-21] [WebCite Cache ID 6aB2Chc4F]
12. Vannette D, Krosnick JA. Answering questions: a comparison of survey satisficing and mindlessness. In: Le A, Ngnoumen CT, Langer EJ, editors. The Wiley Blackwell Handbook of Mindfulness. Chichester, UK: John Wiley & Sons, Ltd; 2004:312-327.
13. Vannette D. How to Avoid Bad Survey Responses?. 2015 Jun 10. URL: <http://www.qualtrics.com/blog/survey-satisficing-protecting-against-unsatisfactory-and-insufficient-responses/> [accessed 2015-07-21] [WebCite Cache ID 6aBhztGxY]
14. Barge S, Gehlbach H. Using the theory of satisficing to evaluate the quality of survey data. Res High Educ 2011 Dec 30;53(2):182-200. [doi: [10.1007/s11162-011-9251-2](https://doi.org/10.1007/s11162-011-9251-2)]
15. Chudoba B. How Much Time are Respondents Willing to Spend on Your Survey?. 2011 Feb 14. URL: [http://blog.surveymonkey.com/blog/2011/02/14/survey\\_completion\\_times/](http://blog.surveymonkey.com/blog/2011/02/14/survey_completion_times/) [WebCite Cache ID 6S9D99kKM]
16. Kapelner A, Chandler D. Preventing Satisficing in Online Surveys: A "Kapcha" to Ensure Higher Quality Data. 2010 Oct 04 Presented at: The World's First Conference on the Future of Distributed Work (CrowdConf 2010); 4 October 2010; San Francisco, CA.
17. Holbrook AL, Anand S, Johnson TP, Cho YI, Shavitt S, Chavez N, et al. Response heaping in interviewer-administered surveys: is it really a form of satisficing? Public Opinion Quarterly 2014 Aug 13;78(3):591-633. [doi: [10.1093/poq/nfu017](https://doi.org/10.1093/poq/nfu017)]
18. Oppenheimer DM, Meyvis T, Davidenko N. Instructional manipulation checks: detecting satisficing to increase statistical power. J Exp Soc Psychol 2009 Jul;45(4):867-872. [doi: [10.1016/j.jesp.2009.03.009](https://doi.org/10.1016/j.jesp.2009.03.009)]
19. Turner G, Sturgis P, Martin D. Can response latencies be used to detect survey satisficing on cognitively demanding questions? J Surv Stat Methodol 2014 Dec 05;3(1):89-108. [doi: [10.1093/jssam/smu022](https://doi.org/10.1093/jssam/smu022)]



20. Hoogendoorn A. A questionnaire design for dependent interviewing that address the problem of cognitive satisficing. *J Off Stat* 2004;20(2):219-232.
21. Carver RP. Rauding Theory predictions of amount comprehended under different purposes and speed reading conditions. *Read Res Q* 1984 Jan;19(2):205-218. [doi: [10.2307/747363](https://doi.org/10.2307/747363)]
22. Dyson M, Haselgrove M. The effects of reading speed and reading patterns on the understanding of text read from screen. *J Res in Reading* 2000 Jun;23(2):210-223. [doi: [10.1111/1467-9817.00115](https://doi.org/10.1111/1467-9817.00115)]
23. Kiwan D, Ahmed A, Pollitt A. The Effects of Time-Induced Stress on Making Inferences During Text Comprehension. 2000 Presented at: 2000 European Conference on Educational Research; 20-23 September 2000; Edinburgh, Scotland.
24. Masson ME. Cognitive processes in skimming stories. *J Exp Psychol-Learn Mem Cogn* 1982;8(5):400-417. [doi: [10.1037/0278-7393.8.5.400](https://doi.org/10.1037/0278-7393.8.5.400)]
25. Zhang C, Conrad F. Speeding in web surveys: the tendency to answer very fast and its association with straightlining. *Surv Res Meth* 2014;8(2):127-135. [doi: [10.18148/srm/2014.v8i2.5453](https://doi.org/10.18148/srm/2014.v8i2.5453)]
26. Malhotra N. Completion time and response order effects in web surveys. *Public Opin Q* 2008 Dec 04;72(5):914-934. [doi: [10.1093/poq/nfn050](https://doi.org/10.1093/poq/nfn050)]
27. Conrad F, Tourangeau R, Couper MP, Zhang C. Interactive Interventions in Web Surveys Can Increase Response Accuracy. 2011 Presented at: Annual meeting of the American Association for Public Opinion Research; 12-15 May 2011; Phoenix, AZ.
28. Wells T, Rao K, Link M, Pierce C. Flagging Speeders in a Multi-Mode (Mobile and Online) Survey. 2012 Presented at: Annual meeting of the Midwest Chapter of the American Association for Public Opinion Research; 16-17 November 2012; Chicago, IL.
29. Bathelt S, Bauknecht J. Effects of speeding on Satisficing in Mixed-Mode Survey. 2011 Presented at: 4th Conference of the European Survey Research Association; 18-22 July 2011; Lausanne, Switzerland.
30. Tourangeau R, Yan T. Sensitive questions in surveys. *Psychol Bull* 2007 Sep;133(5):859-883. [doi: [10.1037/0033-2909.133.5.859](https://doi.org/10.1037/0033-2909.133.5.859)] [Medline: [17723033](https://pubmed.ncbi.nlm.nih.gov/17723033/)]
31. Kaslow RA, Ostrow DG, Detels R, Phair JP, Polk BF, Rinaldo CR. The Multicenter AIDS Cohort Study: rationale, organization, and selected characteristics of the participants. *Am J Epidemiol* 1987 Aug;126(2):310-318. [Medline: [3300281](https://pubmed.ncbi.nlm.nih.gov/3300281/)]
32. Dudley J, Jin S, Hoover D, Metz S, Thackeray R, Chmiel J. The Multicenter AIDS Cohort Study: retention after 9 1/2 years. *Am J Epidemiol* 1995 Aug 1;142(3):323-330. [Medline: [7631636](https://pubmed.ncbi.nlm.nih.gov/7631636/)]
33. MACS Website. URL: <http://www.statepi.jhsph.edu/macs/macs.html> [accessed 2014-08-27] [WebCite Cache ID 6S9DK4fZG]
34. Jones BL, Nagin DS. Advances in group-based trajectory modeling and an SAS procedure for estimating them. *Socio Meth Res* 2007 May 01;35(4):542-571. [doi: [10.1177/0049124106292364](https://doi.org/10.1177/0049124106292364)]
35. Jones BL, Nagin DS, Roeder K. A SAS procedure based on mixture models for estimating developmental trajectories. *Socio Meth Res* 2001 Feb 01;29(3):374-393. [doi: [10.1177/0049124101029003005](https://doi.org/10.1177/0049124101029003005)]
36. Krosnick J, Alwin D. A test of the form-resistant correlation hypothesis: ratings, rankings, and the measurement of values. *Public Opin Q* 1989;52(4):456-538. [doi: [10.1086/269128](https://doi.org/10.1086/269128)]
37. Rao K, Wells T, Luo T. Alert! Marketing Research Association (Fourth Quarter). 2014. Speeders in a multi-mode (mobile and online) survey URL: <http://www.marketingresearch.org/article/speeders-multi-mode-mobile-and-online-survey> [accessed 2016-04-17] [WebCite Cache ID 6gqLphCXM]
38. Holbrook AL, Green MC, Krosnick JA. Telephone versus face-to-face interviewing of national probability samples with long questionnaires. *Public Opin Q* 2003 Mar;67(1):79-125. [doi: [10.1086/346010](https://doi.org/10.1086/346010)]
39. Beckers T, Siegers PA, Kuntz A. Speeders in Online Value Research. 2011 Mar 14 Presented at: General Online Research 11; 14-16 March 2011; Düsseldorf, Germany.
40. Smyth JD. Comparing check-all and forced-choice question formats in web surveys. *Public Opin Q* 2006 Mar 01;70(1):66-77. [doi: [10.1093/poq/nfj007](https://doi.org/10.1093/poq/nfj007)]

## Abbreviations

**ACASI:** audio computer-assisted self interviewing

**AIDS:** acquired immune deficiency syndrome

**CAPI:** computer-assisted personal interviewing

**CATI:** computer-assisted telephone interviewing

**CI:** confidence interval

**HIV:** human immunodeficiency virus

**IQR:** interquartile range

**MACS:** Multicenter AIDS Cohort Study

**OR:** odds ratio



*Edited by G Eysenbach; submitted 15.10.15; peer-reviewed by J Krosnick, K Jones, D Vannette; comments to author 03.04.16; revised version received 13.05.16; accepted 13.06.16; published 08.08.16.*

*Please cite as:*

*Di J, Li Y, Friedman MR, Reddy S, Surkan PJ, Shoptaw S, Plankey M*

*Determining Survey Satisficing of Online Longitudinal Survey Data in the Multicenter AIDS Cohort Study: A Group-Based Trajectory Analysis*

*JMIR Public Health Surveill 2016;2(2):e150*

URL: <http://publichealth.jmir.org/2016/2/e150/>

doi: [10.2196/publichealth.5240](https://doi.org/10.2196/publichealth.5240)

PMID: [27503107](https://pubmed.ncbi.nlm.nih.gov/27503107/)

©Junrui Di, Ying Li, M Reuel Friedman, Susheel Reddy, Pamela J Surkan, Steven Shoptaw, Michael Plankey. Originally published in JMIR Public Health and Surveillance (<http://publichealth.jmir.org>), 08.08.2016. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Public Health and Surveillance, is properly cited. The complete bibliographic information, a link to the original publication on <http://publichealth.jmir.org>, as well as this copyright and license information must be included.

Original Paper

# How Best to Obtain Valid, Verifiable Data Online From Male Couples? Lessons Learned From an eHealth HIV Prevention Intervention for HIV-Negative Male Couples

Jason Mitchell<sup>1\*</sup>, MPH, PhD; Ji-Young Lee<sup>1\*</sup>, MSPH; Rob Stephenson<sup>2</sup>, PhD

<sup>1</sup>Office of Public Health Studies, University of Hawai'i at Manoa, Honolulu, HI, United States

<sup>2</sup>Center for Sexuality and Health Disparities, Department of Health Behavior and Biological Sciences, University of Michigan School of Nursing, Ann Arbor, MI, United States

\*these authors contributed equally

**Corresponding Author:**

Jason Mitchell, MPH, PhD  
Office of Public Health Studies  
University of Hawai'i at Manoa  
1960 East-West Road  
Biomed D 104AA  
Honolulu, HI 96822  
United States  
Phone: 1 808 956 3342  
Fax: 1 808 956 5818  
Email: [jasonmit@hawaii.edu](mailto:jasonmit@hawaii.edu)

**Related Article:**

This is a corrected version. See correction statement: <http://publichealth.jmir.org/2016/2/e155/>

## Abstract

**Background:** As interest increases in the development of eHealth human immunodeficiency virus (HIV)-preventive interventions for gay male couples, Web-based methods must also be developed to help increase the likelihood that couples enrolled and data collected from them represent true unique dyads. Methods to recruit and collect reliable and valid data from both members of a couple are lacking, yet are crucial for uptake of novel sexual health and HIV-prevention eHealth interventions. Methods to describe best practices to recruit male couples using targeted advertisements on Facebook are also lacking in the literature, yet could also help in this uptake.

**Objective:** The objective of our study was to describe challenges and lessons learned from experiences from two phases (developmental phase and online randomized controlled trial [RCT]) of an eHealth HIV-prevention intervention for concordant HIV-negative male couples in terms of (1) recruiting male couples using targeted advertisements on Facebook, (2) validating that data came from two partners of the couple, and (3) verifying that the two partners of the couple are in a relationship with each other.

**Methods:** The developmental phase refined the intervention via in-person focus groups, whereas the pilot-testing phase included an online RCT. For both phases, couples were recruited via targeted Facebook advertisements. Advertisements directed men to a study webpage and screener; once eligible, participants provided consent electronically. A partner referral system was embedded in the consenting process to recruit the relationship partner of the participant. Both men of the couple had to meet all eligibility criteria—individually and as a couple—before they could enroll in the study. Verification of couples' relationships was assessed via the concurrence of predetermined screener items from both partners, done manually in the developmental phase and electronically in the pilot-testing phase. A system of decision rules was developed to assess the validity that data came from two unique partners of a couple.

**Results:** Several important lessons were learned from these experiences, resulting in recommendations for future eHealth studies involving male couples. Use of certain “interests” and types of images (eg, shirtless) in targeted Facebook advertisements should be avoided or used sparingly because these interests and types of images may generate adverse reactions from a broader audience.

Development of a systematic approach with predetermined criteria and parameters to verify male couples' relationships is strongly recommended. Further, researchers are encouraged to develop a system of decision rules to detect and handle suspicious data (eg, suspicious email addresses/names, multiple entries, same IP address used in multiple entries) to help validate the legitimacy of male couples' relationships online.

**Conclusions:** These lessons learned combined with recommendations for future studies aim to help enhance recruitment efforts and the validity and reliability of collecting dyadic data from male couples for novel eHealth HIV-preventive interventions.

(*JMIR Public Health Surveill* 2016;2(2):e152) doi:[10.2196/publichealth.6392](https://doi.org/10.2196/publichealth.6392)

## KEYWORDS

eHealth, male couples; dyadic data; recruitment; validation; verification

## Introduction

Use of the Internet, including social media, has dramatically increased within the past decade. As evidenced in the Pew Research Center's Internet and American Life Project, the prevalence of Internet use among US adults rose from 14% to 87% between 1995 and 2014 [1], and social media use increased from 8% to 76% between 2005 and 2015 [2]. Not only has the Internet allowed individuals and communities to connect and communicate via various multimedia features (eg, chatting, picture sharing, messaging), it has also increasingly been used as a platform to disseminate health information and promote the uptake of healthy behaviors across all scientific disciplines [3-5]. These online environments have created a wide yet complex venue to recruit and collect psychosocial and behavioral data to further health promotion and prevention efforts across different populations.

For example, the Internet has increasingly been used for human immunodeficiency virus (HIV) prevention research with various populations due to the efficiency of targeting and collecting data from specific subpopulations, such as gay, bisexual, and other men who have sex with men (MSM) [6-10]. Despite this efficiency, characteristics of MSM who are recruited online versus in person may differ for some characteristics and not in others. For example, MSM recruited online versus venue-based, time-spaced sampling (ie, offline) did not statistically differ regarding their HIV and sexually transmitted infection (STI) testing patterns, prevalence of HIV and other STIs, and study retention rates up to 24 months [11]. However, online studies with MSM and male couples suggest that Internet-based samples of MSM tend to self-report as being more educated and non-Hispanic white [12,13]. As such, online studies with MSM and male couples may limit the potential for generalizability depending on who participates in the research study.

With respect to prevention, some MSM use the Internet to meet other MSM for sex, friendships, relationships, and for various types of social support [3,14,15]. Although MSM are not the only population to use the Internet for these reasons, research has also noted that Internet use has been associated with having higher numbers of sex partners, greater frequency of condomless anal sex (receptive and insertive), and increased levels of recreational drug use (eg, methamphetamines) among MSM [16,17]. Given these reasons, HIV-prevention efforts that are conducted online may be beneficial to reach some of these MSM in addition to the efforts that are provided in person.

Although methods of recruiting online are more time efficient than recruiting in person, there are challenges related to the reliability and validity of collecting data online from participants [18]. For example, anonymity and lack of direct face-to-face contact with participants prohibit researchers to know who and where data are originating from when it is collected online. In addition, randomly generated responses to survey items may occur in order to help expedite completion of the survey, particularly if an incentive is involved. Monetary incentives have been frequently associated with increased participant misrepresentation for eligibility [19,20] and multiple data entry [20,21] by individuals who wish to increase the amount of incentives and/or probabilities of winning. Survey response rates may also decrease over time, which could threaten the validity of generalizing the results to a broader audience. Bauermeister and colleagues [22] examined how invalid data collected online may influence statistical relationships, decrease statistical power, and increase the likelihood of biased conclusions. They did not recommend using a conservative approach and excluding suspicious entries because valid data could accidentally be removed in the process. Instead, they recommended developing both pre and post hoc decisions to handle all data, such as grouping entries into different categories (eg, valid, suspicious, and invalid) and obtaining a population list of all eligible participants for verification processes, respectively [22].

These issues of reliability and validity of online data become more complex when conducting online research with dyads (eg, gay male couples). By nature, handling dyadic data requires at least double the amount of time and effort than that required of individual-level data. The criteria used to determine the reliability and validity of an individual participant must be assessed not only twice, but must also be crosschecked with that of that participants' partner's data to (1) validate the data are coming from two unique individuals and (2) verify that these two unique individuals are in a relationship together. The eligibility criteria used in couples Web-based research is also inherently more complex due to the need to ensure that both individual- and couple-level eligibility requirements are met. However, the anonymity that the Internet warrants makes it easier for participants to pose as a fraudulent "couple" by pretending to be both partners in the relationship. If incentives are used, then this further increases the need to ensure that dyadic data collected online are valid and reliable because some couples' motivation to become eligible may be less than honest. For these reasons and due to the increased use of the Internet

among adults, methods are needed to help ensure that dyadic data collected online from couples are valid and verifiable of their relationships.

In recent years, attention toward male couples for HIV prevention has increased because many MSM in the United States—up to 67%—acquire HIV while in a same-sex relationship [23,24]. In addition, few preventive interventions approved by the Centers for Disease Control and Prevention currently exist for this population [25-28]. Despite how common Internet use is among MSM and its efficiency to enhance recruitment and data collection efforts for prevention research, few studies have fully harnessed the capabilities of eHealth (ie, Web-based research and/or preventive programs hosted online) to help advance HIV-prevention efforts for male couples [29]. Moreover, the majority of HIV-preventive interventions for MSM have focused on behavior change at the individual level [30-33], thereby emphasizing the need to not only develop eHealth HIV-preventive interventions for male couples, but to also develop methods to maximize recruitment efforts and verification of valid data from both members of the couple (ie, dyadic data) online. Methods to recruit and collect reliable and valid dyadic data online are lacking, yet are crucial for development and uptake of future eHealth HIV-preventive interventions for male couples.

To help fill this critical gap, this paper aims to provide methodological recommendations in three distinct areas to (1) maximize efficiency and accuracy of using targeted, online Facebook advertisements to recruit male couples; (2) facilitate the collection of *valid* dyadic data; and (3) ensure verification that dyadic data collected are representative of two males who are in a relationship together (ie, male couple). To accomplish these aims, data and related experiences captured from the development and pilot testing of an eHealth HIV-prevention intervention with concordant HIV-negative male couples in the United States are used. The methods used and lessons learned from these experiences provide concrete suggestions and recommended safeguards that may benefit future eHealth endeavors targeting similar populations of male couples.

## Methods

### Procedure Overview

The University of Miami Institutional Review Board approved all study procedures. For both the development and pilot testing of the eHealth HIV-prevention intervention, interested men who clicked on the Facebook advertisement were directed to the study webpage and an eligibility screener via SurveyGizmo, a Health Insurance Portability and Accountability Act-compliant Web-based survey tool and database server. Once eligible, participants provided consent electronically. A partner referral system was embedded in the consent process to recruit the relationship partner of the index participant. Both men of the couple had to meet all eligibility criteria and be deemed valid—as an individual and as a couple—before they could enroll into either the development or pilot-testing portion of the study; the development portion included both partners of the couple participating in one in-person focus group, whereas the pilot testing of the intervention included a Web-based

randomized controlled trial (RCT) with couples participating online.

### Targeted Recruitment via Facebook

Male couples were recruited via targeted advertisements placed on Facebook. All advertisements targeted potential research participants who were male, living in the United States, at least 18 years of age, interested in men, and had a relationship status of either being married, engaged, in a relationship, domestic partnership, or civil union. Each advertisement included a picture of a male couple with a brief title, message, and a Web link to the study eligibility screener. In total, 13 advertisement campaigns were conducted for the focus group, development phase, and RCT pilot-testing phase of the project, and each Facebook advertisement campaign lasted for 72 hours. The campaigns used to recruit for the focus groups occurred between March 2015 and May 2015, whereas the campaigns used to recruit for the online RCT occurred between October 2015 and March 2016. Following Facebook's guidelines and word limits, all advertisements contained titles similar to "In a relationship?" with a brief message such as "Male couples wanted to try out a cool, new online health & relationship program. Earn \$\$!" Each campaign's total cost ranged from US \$499.97 to US \$3997.62 depending on performance, audience, and placement of the advertisements (eg, desktop newsfeed, mobile newsfeed). Per campaign, the advertisements resulted in a mean of 110,478 people reached and a mean of 3534 clicks to the study website. Each targeted Facebook advertisement campaign provided useful metrics about how well the targeted advertisement performed, including (1) the total number of people reached, (2) number of people who clicked on the advertisement, (3) how relevant the targeted audience thought the advertisement pertained to them (ie, relevancy score ranging from 0 to 10 with 10 being most relevant), and (4) negative and positive feedback about the advertisement obtained from those who were shown the advertisement while using Facebook.

### Eligibility Criteria

Both members of the male couple had to meet the following eligibility criteria to participate in the focus group or RCT of the intervention project: (1) self-report as male, (2) be at least 18 years of age, (3) be in a current sexual relationship with a main relationship partner for at least 6 months, (4) self-report as HIV-negative, (5) practice condomless anal sex with the main relationship partner for at least 6 months, (6) self-report no recent history of intimate partner violence or coercion within the previous year, (7) have not established a sexual agreement in the relationship, and (8) own a mobile phone and have an alternate method to access the Internet (eg, computer).

### Procedures Used to Verify Couples' Relationships and Validity of Dyadic Data Collected

Both individual- and couple-level criteria were used to deem couples' eligibility to participate in the study. Based on our prior work with male couples [34,35], we developed a system of decision rules to assess the verification of a couples' relationship based on both partners' responses to screener items (ie, to verify that a couple is in a relationship). Verification of couples' relationships was based on data collected from both



members of the couple and whether they concurred (ie, agreed) on certain predetermined eligibility screener items (see Figure 1), including (1) matching their relationship length within a margin of 1 month, (2) exact matching of each other’s birthday months, (3) matching each other’s ages within a margin of one year, and (4) exact matching of at least one of their two contact information options (eg, phone number, email address). All couples had to concur on all four items to be considered eligible. Each participant and his partner’s data were crosschecked to match and concur on all four of the predetermined screener items. Suspicious cases or cases in which couples did not concur on one or two items were assessed a second time after calling or emailing the potential participant to verify their corresponding partners’ information. Verification of couples’ relationships was done manually by entering data into a secure database during the developmental focus group phase; for the pilot RCT, this verification step was done electronically via an embedded Web-based system that contained an algorithm.

For both phases of the intervention project, validation of both partners’ data collected was assessed on a case-by-case basis.

For the development phase, multiple entries of the same Internet Protocol (IP) address were examined to validate whether or not one or both partners were unique individuals. For example, each entry of the eligibility screener had an IP address associated with it. Some IP addresses were associated with several entries, potentially suggesting that individuals had attempted to complete the eligibility screener multiple times as either the same person or pretending to be different people. Eligibility screener entries with the same IP address were flagged and noted for further investigation by using the other validation markers.

Because the online pilot RCT aimed to recruit a large number of male couples (compared to the focus groups), an electronic algorithm was created to electronically match both partners as a couple based on predetermined eligibility criteria and verification rules. For the online RCT phase, suspicious names and/or email addresses, duplicate entries (ie, same two email addresses used for two different sets of couples), and similar back-to-back screener entries were examined in addition to the validation markers used in the developmental focus group phase.

**Figure 1.** An example of some of the items used in the verification of couples’ relationships via manual analysis of the eligibility data collected from both partners online. Predetermined variables with decision rules were used to verify whether two partners were in a relationship with each other. Note: data represented in this figure are fictitious.

ID	bday mo.	p2 bday mo.	age	p2 age	LTR	SA	partner email	email	cell phone	p2 cell phone	IP address	Time start	Time end	Notes
149	October	April	20	25	1-3 yrs	No	Matthew.s2@gm	travlee@gmail.co	need	need	4.71.214.87	4/18/2015 0:24:0	4/18/2015 0:28:0	
76	February	December	45	54	6-10 yrs	Yes	jcollins@icloud.c	mwellington@gm	3085492146	9657475514	174.61.16.131	4/8/2015 21:50:0	4/8/2015 21:55:0	ELIGIBLE
108	December	February	54	45	6-10 yrs	No	mwellington@gm	jcollins@icloud.c	9657475514	3085492146	174.71.26.131	4/15/2015 14:57:0	4/15/2015 15:01:0	COUPLE
168	August	March	44	45	>10 yrs	No	asimmons590@i	ppaterson78@gr	6485669718	9458523645	73.464.184.10	4/18/2015 9:53:0	4/18/2015 9:56:0	ELIGIBLE, NOT MEET 1 PARAMETER
177	March	August	45	44	>10 yrs	Yes	ppaterson78@gr	ajs65@pitt.edu	9458523645	6485669718	73.204.854.1C	4/18/2015 16:11:0	4/18/2015 16:14:0	MEET 1 PARAMETER
195	April	February	50	53	>10 yrs	Yes	carlguilt23@ho	tmburrs@gmail.c	5687458213	3694587561	73.179.945.96	4/18/2015 22:36:0	4/18/2015 22:40:0	INELIGIBLE, SA & did not provide partner contact info
291	February	April	53	50	>10 yrs	Yes	tmburrs@gmail.c	carlguilt23@ho						
107	March	July	43	33	1-3 yrs	No	wwickham938@	e.stanford68@hc	8751469782	need, sent	50.852.158.1	4/15/2015 14:42:0	4/15/2015 14:58:0	FRAUD- SAME RESPONSES FOR 211 & 212
213	October	March	32	43	3-6 yrs	No	e.stanford68@hc	wwickham938@	7862549853	8751469782	50.192.624.1	4/19/2015 9:53:0	4/19/2015 9:59:0	

## Results

Lessons learned from this intervention project are an example of the challenges associated with collecting and verifying valid dyadic data from male couples online. The lessons learned from this project are not exhaustive, yet provide insights and suggestions to help improve these efforts, particularly with the increased attention toward developing eHealth and mHealth preventive interventions for male couples [36-38]. Specifically, lessons learned from this project entail (1) advantages and disadvantages of using certain selection criteria to target advertisements on Facebook, (2) selection of images that resonate with the targeted audience, (3) use of an algorithm versus manual input of data to verify male couples’ relationships via predetermined decision rules, and (4) monitoring of dyadic

data collected to assess validity of unique responses obtained from both partners of the male couple.

### Targeted Advertisements for Male Couples on Facebook

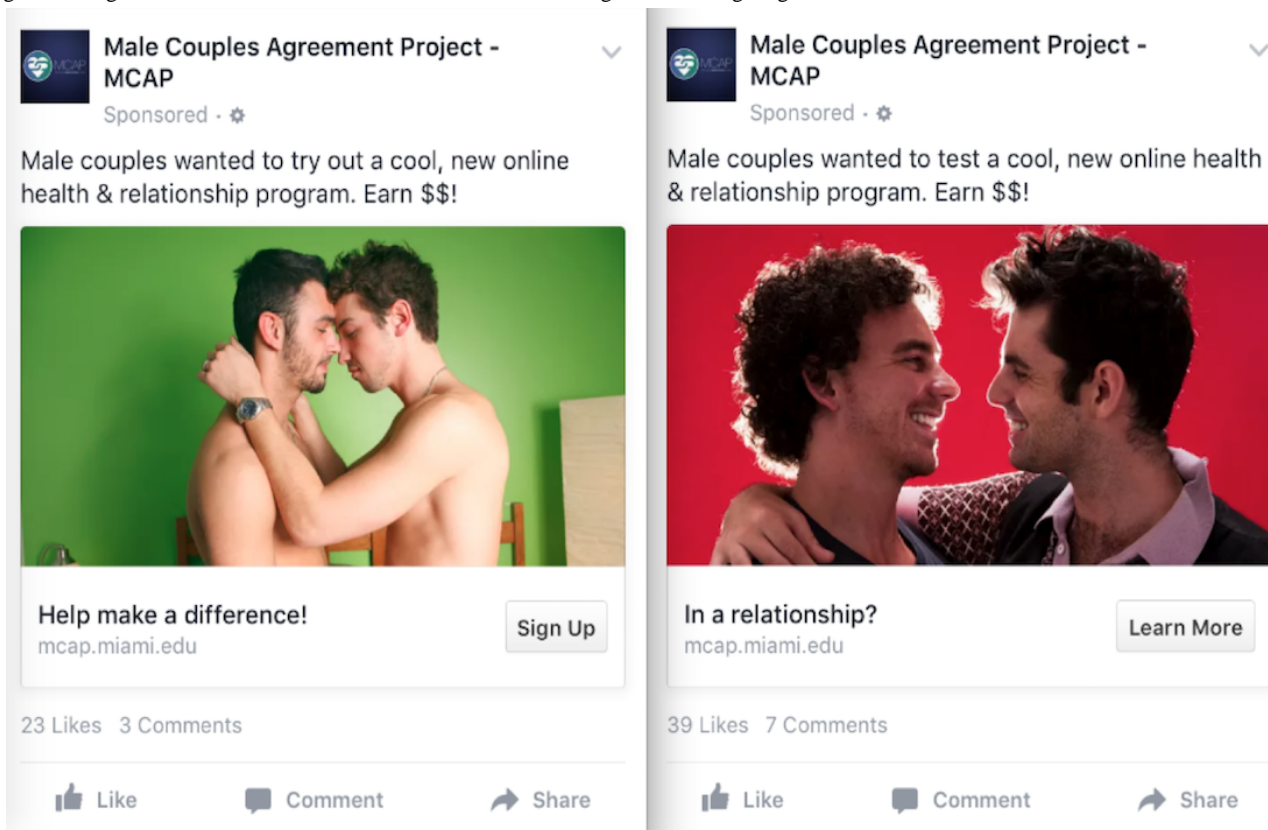
Some images used in the advertisements received higher relevancy scores than others and received less negative feedback from supportive individuals who were not lesbian, gay, bisexual, and transgender (LGBT). Negative feedback was received in three forms: (1) public comments posted on the Facebook advertisements and study Facebook community page, (2) messages sent privately to the study Facebook community page, and (3) voicemail. Figure 2 provides examples of the negative feedback received from some of the images used for the targeted advertisements placed on Facebook.



**Figure 2.** Examples of negative feedback received for some of the targeted advertisements used on Facebook.



**Figure 3.** Example of Facebook advertisements before and after modifications. Image on the left was used before negative feedback was received. Image on the right was used after modifications were made to the images and the targeting criteria.



After some investigation and consultation with Facebook (private telephone communication with Facebook, January 13, 2016), interests that were used to target Facebook members, such as the LGBT community or gay news, included members who had positive views of this interest as well as those who had negative views of this same interest. Because the negative-view individuals were also targeted, this resulted in a flood of

backlash in the form of unsupportive and homophobic comments posted on the study Facebook webpage, study team’s voicemail, and as comments left in association with the targeted advertisement. After these experiences, targeting criteria were adjusted for the remainder of the advertisement campaigns by actively excluding Facebook members who had the following information on or associated with their profiles: interested in

women or both men and women; had a relationship status as single, unspecified, open, complicated, separated, divorced, or widowed; and had moderate, conservative, and very conservative political views. Targeted advertisements were also changed by using less overtly sexual images of male couples (eg, clothed male couple slightly embracing). [Figure 3](#) provides examples of Facebook advertisements that were used before and after these modifications were made.

### Verification of Couples' Relationships

A systematic manual (ie, color-coded) approach was used during the project's developmental phase to verify whether both partners of the male couple were in a relationship (after both men had electronically consented) using predetermined eligibility screener items. This post hoc comparative method of the dyadic data collected from the eligibility screener enabled the characterization of couples into one of three groups: (1) eligible and verified couples who met all predetermined decision rules, (2) eligible but unverified couples because only two or three of the predetermined decision rules were met, and (3) ineligible couples. This systematic manual method illuminated the strengths and limitations of this approach to verify couples' relationships. Although manually matching entries allowed for greater control over the assessment of couples' relationships, this method proved inefficient in terms of time and resources.

Based on this experience, the amount of time required to verify couples' relationships with this approach emphasized the necessity to electronically automate this process, particularly if data were to be collected online from large samples of male couples such as during an eHealth intervention. As such, a Web-based electronic algorithm was created to automatically verify couples' relationships based on data collected from both members of the couple and whether they concurred on predetermined eligibility screener items. The algorithm was developed based on the results and experiences obtained during the developmental focus group phase. However, the algorithm had an important limitation. The system would report couples as partially matched when their contact information mismatched

(eg, one partner reported his partners' email address different from what his partner reported as his actual email address). This mismatch required a manual verification of the couples' relationship by contacting the potential participants to further assess the couples' relationship and their eligibility to enroll into the project.

### Validation of Dyadic Data Collected

For time- and resource-efficiency reasons, validation of couples' data (ie, data came from two unique individuals) occurred after the algorithm electronically matched and verified that a couple was in a legitimate relationship. This approach helped to expedite the recruitment process by filtering out ineligible couples and couples whose relationships were not verified. Verified couples could then be validated. Validation of couples' data was still required because some instances noted that the same IP addresses were being used in multiple screener entries. In some of these sets of entries with the same IP addresses, it was clear that the respondent(s) attempted the screener multiple times to determine the eligibility criteria to enroll in the study. Some individuals also created generic email addresses to appear as a couple for inclusion into the study. Because of these instances, the awareness of fraudulent yet eligible entries increased, thus warranting the creation of additional safeguards for a post hoc assessment of verified screener entries. Specifically, similar and/or suspicious email addresses (eg, "kylemcap1@gmail.com" and reporting partner's to be "kylemcap2@gmail.com") or names (eg, Melissa Wise) were noted and flagged. Also noted were similar yet different screener back-to-back entries to determine eligibility, same IP addresses used in multiple screener entries, and duplicate entries (ie, same two emails used for two different sets of couples). For these cases, the individuals and their "partners" were called and asked to verify, via phone, specific questions to assess their validity. [Table 1](#) illustrates the difference between the manual and electronic methods of matching couples (verification of couples' relationships) as well as examples of the post hoc assessment of screener entries (validation that data were from two unique individuals).

**Table 1.** Summary of the verification and validation of couples' relationships and data for each of phase of the project.

Information used to verify and validate couples' relationships	Developmental focus group phase	Pilot-testing randomized controlled trial phase
Number of screener items	17	17
Number of items used to determine eligibility	10	11
Predetermined items used to assess verification of relationship via comparison of self-reports by both partners of the couple	Individual's birthday month; partner's birthday month; individual's age; partner's age; relationship length; sexual agreement; individual's email; partner's email; individual's phone number; partner's phone number	Post hoc assessment
Method used to match and crosscheck predetermined items for relationship verification	Manual	Electronic via algorithm
Post hoc assessment of screener entries for validation purposes <sup>a</sup>	Similar/suspicious email addresses and names (eg, Kylemcap1@gmail.com and kylemcap2@gmail.com; Melissa Wise); back-to-back completions: time-stamped as 20-30 minutes apart (duration of survey); similarity of responses: consistent extreme/haphazard responses throughout survey; same IP addresses used in multiple entries: same IP address appearing in multiple screener entries with different responses (potentially to determine and pass eligibility criteria); duplicate entries: same two emails used for two different sets of couple IDs (eg, same two emails used for two different IDs: 2452EW and 2864OD)	

<sup>a</sup> Experience suggested the continued need for post hoc assessment when using an electronic method and/or dealing with large sample sizes.

## Discussion

Several important lessons were learned about recruiting and collecting data from male couples from this project, resulting in recommendations for future Web-based studies specific to male couples. Because many MSM and/or male couples used the study Facebook webpage to ask questions or share stories/details, it was important to provide a friendly and supportive environment for these men to connect and communicate, free from public hostile comments.

For placement of targeted advertisements on Facebook, certain "interests" (eg, LGBT news) and images (ie, shirtless) should be avoided or used sparingly because they may generate adverse reactions from a broader audience. Although all images used in the advertisements were appropriate, some images may have been too explicit or hypersexualized for those outside the targeted audience, and especially to those who explicitly have a negative attitude toward the LGBT community. Researchers may also want to consider contacting established Facebook community groups (eg, Gay Miami Beach) who may have members of the targeted population to help promote the study. With respect to future work, researchers should consider how other characteristics of partnered MSM and male couples may influence the likelihood of them responding to targeted advertisements placed on Facebook. For instance, some images and recruitment text may resonate better with certain subgroups of male couples (eg, younger vs older male couples, ethnicity/race) and not others. Another consideration includes the target population's computer and research literacy. As such, formative work is recommended to investigate which images, text, and placement of the advertisement (eg, desktop newsfeed, mobile newsfeed) would best help enhance the target populations' response to clicking on the Facebook advertisement.

Experiences associated with screening potential male couples and verifying their relationships reinforces the need to develop

and use a systematic approach that includes predetermined criteria with decision rules. These parameters will help enhance the efficiency as well as the reliability of collecting data from both members of the male couple. However, although time efficient, algorithms to electronically verify couples' relationships may be imperfect and unable to capture certain nuances of their relationships. For example, two partners may define their relationship length differently if they had a break or period of separation occur during their relationship. Due to the discordance in their self-reports about relationship length, the predetermined decision rule associated with this questionnaire item would result in classifying this particular couples' relationship as unverified. Based on the algorithm, this couple would then be deemed as unverified because of this one criterion they disagreed on. Researchers should consider which criteria they plan to use with corresponding decision rules because these decisions may affect the eligibility and verification of couples' relationships in their Web-based projects that are specific to male couples.

It also is important to note that validation of dyadic data should not be assessed in isolation. Validation of dyadic data collected online should be based on examining multiple, if not all, predetermined criteria. Examining only one criterion (eg, same IP addresses, suspicious emails) may lead to disqualifying a potentially eligible verified couple. For instance, both partners of a male couple may live together and use the same computer, which would report them having the same IP address when completing a Web-based questionnaire. Email addresses that appear deceiving or misrepresentative of a potential participant warrant further investigation. For instance, an email address of "heather1@gmail.com" could be flagged as suspicious because the first name was Heather, a name typically associated with being a female. However, further investigation of this participant could reveal that his name was "Heath" and not "Heather."

The Internet is a useful platform to target and enroll at-risk MSM and male couples from a variety of locales. However, it is evident that recruiting and collecting valid and reliable dyadic

data online is more complex than that at the individual level. The varying results of the Facebook advertisement campaigns to recruit male couples further stresses the need to monitor posted Facebook advertisements to help minimize the receipt of negative feedback and that the target audience is being reached. Developing a systematic approach with predetermined criteria and parameters is strongly recommended to verify male couples' relationships as efficiently as possible. Researchers are encouraged to develop a system of decision rules to detect

and handle suspicious data to help validate the legitimacy that the dyadic data are coming from two unique partnered individuals. With the decision rules and criteria set, researchers should be cautious in observing multiple criteria as a whole, rather than in isolation. These lessons learned combined with recommendations for future studies may help other researchers enhance recruitment efforts and the validity and reliability of collecting dyadic data from male couples for novel preventive eHealth interventions specific to this population.

---

## Acknowledgments

Special thanks are extended to the participants for their time and effort in participating in this study. Without their participation, improving these methods would not have been possible. This work was supported by the National Institute of Mental Health under Grant R34 MH102098 (PI Mitchell J).

---

## Conflicts of Interest

None declared.

---

## References

1. Pew Research Center. 2015. Internet use over time URL: <http://www.pewinternet.org/data-trend/internet-use/internet-use-over-time/> [accessed 2016-07-06] [WebCite Cache ID 6inqSBRt1]
2. Pew Research Center. 2015. Social media use over time URL: <http://www.pewinternet.org/data-trend/social-media/social-media-use-all-users/> [accessed 2016-07-06] [WebCite Cache ID 6inqYVuA0]
3. Kubicek K, Carpineto J, McDavitt B, Weiss G, Kipke MD. Use and perceptions of the internet for sexual information and partners: a study of young men who have sex with men. *Arch Sex Behav* 2011 Aug;40(4):803-816 [FREE Full text] [doi: [10.1007/s10508-010-9666-4](https://doi.org/10.1007/s10508-010-9666-4)] [Medline: [20809373](https://pubmed.ncbi.nlm.nih.gov/20809373/)]
4. Magee JC, Bigelow L, Dehaan S, Mustanski BS. Sexual health information seeking online: a mixed-methods study among lesbian, gay, bisexual, and transgender young people. *Health Educ Behav* 2012 Jun;39(3):276-289. [doi: [10.1177/1090198111401384](https://doi.org/10.1177/1090198111401384)] [Medline: [21490310](https://pubmed.ncbi.nlm.nih.gov/21490310/)]
5. Minichiello V, Rahman S, Dune T, Scott J, Dowsett G. E-health: potential benefits and challenges in providing and accessing sexual health services. *BMC Public Health* 2013;13:790 [FREE Full text] [doi: [10.1186/1471-2458-13-790](https://doi.org/10.1186/1471-2458-13-790)] [Medline: [23987137](https://pubmed.ncbi.nlm.nih.gov/23987137/)]
6. Grossman CI, Forsyth A, Purcell DW, Allison S, Toledo C, Gordon CM. Advancing novel HIV prevention intervention research with MSM--meeting report. *Public Health Rep* 2011;126(4):472-479 [FREE Full text] [Medline: [21800742](https://pubmed.ncbi.nlm.nih.gov/21800742/)]
7. Rosser BR, Wilkerson JM, Smolenski DJ, Oakes JM, Konstan J, Horvath KJ, et al. The future of Internet-based HIV prevention: a report on key findings from the Men's INternet (MINTS-I, II) Sex Studies. *AIDS Behav* 2011 Apr;15 Suppl 1:S91-S100 [FREE Full text] [doi: [10.1007/s10461-011-9910-5](https://doi.org/10.1007/s10461-011-9910-5)] [Medline: [21360127](https://pubmed.ncbi.nlm.nih.gov/21360127/)]
8. Strömdahl S, Lu X, Bengtsson L, Liljeros F, Thorson A. Implementation of Web-based respondent driven sampling among men who have sex with men in Sweden. *PLoS One* 2015;10(10):e0138599 [FREE Full text] [doi: [10.1371/journal.pone.0138599](https://doi.org/10.1371/journal.pone.0138599)] [Medline: [26426802](https://pubmed.ncbi.nlm.nih.gov/26426802/)]
9. Thériault N, Bi P, Hiller JE, Nor M. Use of web 2.0 to recruit Australian gay men to an online HIV/AIDS survey. *J Med Internet Res* 2012;14(6):e149 [FREE Full text] [doi: [10.2196/jmir.1819](https://doi.org/10.2196/jmir.1819)] [Medline: [23128646](https://pubmed.ncbi.nlm.nih.gov/23128646/)]
10. Zou H, Wu Z, Yu J, Li M, Ablimit M, Li F, et al. Internet-facilitated, voluntary counseling and testing (VCT) clinic-based HIV testing among men who have sex with men in China. *PLoS One* 2013;8(2):e51919 [FREE Full text] [doi: [10.1371/journal.pone.0051919](https://doi.org/10.1371/journal.pone.0051919)] [Medline: [23418417](https://pubmed.ncbi.nlm.nih.gov/23418417/)]
11. Hernandez-Romieu AC, Sullivan PS, Sanchez TH, Kelley CF, Peterson JL, Del RC, 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](https://doi.org/10.2196/resprot.3342)] [Medline: [25048694](https://pubmed.ncbi.nlm.nih.gov/25048694/)]
12. Mitchell JW, Sophus AI, Lee J, Petroll AE. Anal douche practices and willingness to use a rectal microbicide enema for HIV prevention and associated factors among an internet sample of HIV-negative and HIV-discordant male couples in the US. *AIDS Behav* 2015 Nov 23:1250-1254. [doi: [10.1007/s10461-015-1250-4](https://doi.org/10.1007/s10461-015-1250-4)] [Medline: [26597502](https://pubmed.ncbi.nlm.nih.gov/26597502/)]
13. Sanchez T, Zlotorzynska M, Sineath C, Kahle E, Sullivan P. The annual American Men's Internet Survey of Behaviors of Men Who have Sex with Men in the United States: 2014 key indicators report. *JMIR Public Health Surveill* 2016;2(1):e23 [FREE Full text] [doi: [10.2196/publichealth.5476](https://doi.org/10.2196/publichealth.5476)] [Medline: [27244770](https://pubmed.ncbi.nlm.nih.gov/27244770/)]
14. Al-Tayyib AA, McFarlane M, Kachur R, Rietmeijer CA. Finding sex partners on the internet: what is the risk for sexually transmitted infections? *Sex Transm Infect* 2009 Jun;85(3):216-220. [doi: [10.1136/sti.2008.032631](https://doi.org/10.1136/sti.2008.032631)] [Medline: [19098059](https://pubmed.ncbi.nlm.nih.gov/19098059/)]



15. Kakietek J, Sullivan PS, Heffelfinger JD. You've got male: internet use, rural residence, and risky sex in men who have sex with men recruited in 12 U.S. cities. *AIDS Educ Prev* 2011 Apr;23(2):118-127 [[FREE Full text](#)] [doi: [10.1521/aeap.2011.23.2.118](https://doi.org/10.1521/aeap.2011.23.2.118)] [Medline: [21517661](https://pubmed.ncbi.nlm.nih.gov/21517661/)]
16. Benotsch EG, Kalichman S, Cage M. Men who have met sex partners via the Internet: prevalence, predictors, and implications for HIV prevention. *Arch Sex Behav* 2002 Apr;31(2):177-183. [Medline: [11974643](https://pubmed.ncbi.nlm.nih.gov/11974643/)]
17. Ko N, Tseng P, Huang Y, Chen Y, Hsu S. Seeking sex partners through the internet and mobile phone applications among men who have sex with men in Taiwan. *AIDS Care* 2016 Jul;28(7):927-931. [doi: [10.1080/09540121.2015.1131969](https://doi.org/10.1080/09540121.2015.1131969)] [Medline: [26754350](https://pubmed.ncbi.nlm.nih.gov/26754350/)]
18. Pequegnat W, Rosser BR, Bowen AM, Bull SS, DiClemente RJ, Bocking WO, et al. Conducting Internet-based HIV/STD prevention survey research: considerations in design and evaluation. *AIDS Behav* 2007 Jul;11(4):505-521. [doi: [10.1007/s10461-006-9172-9](https://doi.org/10.1007/s10461-006-9172-9)] [Medline: [17053853](https://pubmed.ncbi.nlm.nih.gov/17053853/)]
19. Kramer J, Rubin A, Coster W, Helmut E, Hermos J, Rosenbloom D, et al. Strategies to address participant misrepresentation for eligibility in Web-based research. *Int J Methods Psychiatr Res* 2014 Mar;23(1):120-129 [[FREE Full text](#)] [doi: [10.1002/mpr.1415](https://doi.org/10.1002/mpr.1415)] [Medline: [24431134](https://pubmed.ncbi.nlm.nih.gov/24431134/)]
20. Quach S, Pereira JA, Russell ML, Wormsbecker AE, Ramsay H, Crowe L, et al. The good, bad, and ugly of online recruitment of parents for health-related focus groups: lessons learned. *J Med Internet Res* 2013 Jan;15(11):e250 [[FREE Full text](#)] [doi: [10.2196/jmir.2829](https://doi.org/10.2196/jmir.2829)] [Medline: [24231040](https://pubmed.ncbi.nlm.nih.gov/24231040/)]
21. Morabia A, Zheng Y. On the influence of a raffle upon responses to an urban transportation survey in New York City. *Int J Public Health* 2009;54(1):31-34. [doi: [10.1007/s00038-008-8118-6](https://doi.org/10.1007/s00038-008-8118-6)] [Medline: [19190981](https://pubmed.ncbi.nlm.nih.gov/19190981/)]
22. Bauermeister J, Pingel E, Zimmerman M, Couper M, Carballo-Diéguez A, Strecher VJ. Data quality in web-based HIV/AIDS research: handling invalid and suspicious data. *Field Methods* 2012 Aug 1;24(3):272-291 [[FREE Full text](#)] [doi: [10.1177/1525822X12443097](https://doi.org/10.1177/1525822X12443097)] [Medline: [23180978](https://pubmed.ncbi.nlm.nih.gov/23180978/)]
23. Goodreau SM, Carnegie NB, Vittinghoff E, Lama JR, Sanchez J, Grinsztejn B, et al. What drives the US and Peruvian HIV epidemics in men who have sex with men (MSM)? *PLoS One* 2012;7(11):e50522 [[FREE Full text](#)] [doi: [10.1371/journal.pone.0050522](https://doi.org/10.1371/journal.pone.0050522)] [Medline: [23209768](https://pubmed.ncbi.nlm.nih.gov/23209768/)]
24. Sullivan PS, Salazar L, Buchbinder S, Sanchez TH. Estimating the proportion of HIV transmissions from main sex partners among men who have sex with men in five US cities. *AIDS* 2009 Jun 1;23(9):1153-1162. [doi: [10.1097/QAD.0b013e32832baa34](https://doi.org/10.1097/QAD.0b013e32832baa34)] [Medline: [19417579](https://pubmed.ncbi.nlm.nih.gov/19417579/)]
25. Burton J, Darbes LA, Operario D. Couples-focused behavioral interventions for prevention of HIV: systematic review of the state of evidence. *AIDS Behav* 2010 Feb;14(1):1-10 [[FREE Full text](#)] [doi: [10.1007/s10461-008-9471-4](https://doi.org/10.1007/s10461-008-9471-4)] [Medline: [18843530](https://pubmed.ncbi.nlm.nih.gov/18843530/)]
26. Grossman CI, Forsyth A, Purcell DW, Allison S, Toledo C, Gordon CM. Advancing novel HIV prevention intervention research with MSM--meeting report. *Public Health Rep* 2011;126(4):472-479 [[FREE Full text](#)] [Medline: [21800742](https://pubmed.ncbi.nlm.nih.gov/21800742/)]
27. Herbst JH, Beeker C, Mathew A, McNally T, Passin WF, Kay LS, Task Force on Community Preventive Services. The effectiveness of individual-, group-, and community-level HIV behavioral risk-reduction interventions for adult men who have sex with men: a systematic review. *Am J Prev Med* 2007 Apr;32(4 Suppl):S38-S67. [doi: [10.1016/j.amepre.2006.12.006](https://doi.org/10.1016/j.amepre.2006.12.006)] [Medline: [17386336](https://pubmed.ncbi.nlm.nih.gov/17386336/)]
28. El-Bassel N, Gilbert L, Witte S, Wu E, Hunt T, Remien RH. Couple-based HIV prevention in the United States: advantages, gaps, and future directions. *J Acquir Immune Defic Syndr* 2010 Dec;55 Suppl 2:S98-S101. [doi: [10.1097/QAI.0b013e3181fbf407](https://doi.org/10.1097/QAI.0b013e3181fbf407)] [Medline: [21406997](https://pubmed.ncbi.nlm.nih.gov/21406997/)]
29. Wu E, El-Bassel N, Donald ML, Fontaine Y, Hess L. Adaptation of a couple-based HIV intervention for methamphetamine-involved African American men who have sex with men. *Open AIDS J* 2010;4:123-131 [[FREE Full text](#)] [doi: [10.2174/1874613601004030123](https://doi.org/10.2174/1874613601004030123)] [Medline: [20657720](https://pubmed.ncbi.nlm.nih.gov/20657720/)]
30. Hirshfield S, Chiasson MA, Joseph H, Scheinmann R, Johnson WD, Remien RH, et al. An online randomized controlled trial evaluating HIV prevention digital media interventions for men who have sex with men. *PLoS One* 2012;7(10):e46252 [[FREE Full text](#)] [doi: [10.1371/journal.pone.0046252](https://doi.org/10.1371/journal.pone.0046252)] [Medline: [23071551](https://pubmed.ncbi.nlm.nih.gov/23071551/)]
31. Wei C, Herrick A, Raymond HF, Anglemeyer A, Gerbase A, Noar SM. Social marketing interventions to increase HIV/STI testing uptake among men who have sex with men and male-to-female transgender women. *Cochrane Database Syst Rev* 2011(9):CD009337. [doi: [10.1002/14651858.CD009337](https://doi.org/10.1002/14651858.CD009337)] [Medline: [21901734](https://pubmed.ncbi.nlm.nih.gov/21901734/)]
32. Ybarra ML, Bull SS, Prescott TL, Birungi R. Acceptability and feasibility of CyberSenga: an Internet-based HIV-prevention program for adolescents in Mbarara, Uganda. *AIDS Care* 2014 Apr;26(4):441-447 [[FREE Full text](#)] [doi: [10.1080/09540121.2013.841837](https://doi.org/10.1080/09540121.2013.841837)] [Medline: [24093828](https://pubmed.ncbi.nlm.nih.gov/24093828/)]
33. Young SD, Zhao M, Teiu K, Kwok J, Gill H, Gill N. A social-media based HIV prevention intervention using peer leaders. *J Consum Health Internet* 2013 Oct 1;17(4):353-361 [[FREE Full text](#)] [doi: [10.1080/15398285.2013.833445](https://doi.org/10.1080/15398285.2013.833445)] [Medline: [24526928](https://pubmed.ncbi.nlm.nih.gov/24526928/)]
34. Mitchell JW. Characteristics and allowed behaviors of gay male couples' sexual agreements. *J Sex Res* 2014;51(3):316-328 [[FREE Full text](#)] [doi: [10.1080/00224499.2012.727915](https://doi.org/10.1080/00224499.2012.727915)] [Medline: [23514544](https://pubmed.ncbi.nlm.nih.gov/23514544/)]



35. Mitchell JW, Lee JY, Woodyatt C, Bauermeister J, Sullivan P, Stephenson R. Perceived challenges and rewards of forming a sexual agreement among HIV-negative male couples. *Arch Sex Behav* 2016 Aug;45(6):1525-1534. [doi: [10.1007/s10508-016-0701-y](https://doi.org/10.1007/s10508-016-0701-y)] [Medline: [26964794](#)]
36. Cordova D, Bauermeister JA, Fessler K, Delva J, Nelson A, Nurenberg R, et al. A community-engaged approach to developing an mHealth HIV/STI and drug abuse preventive intervention for primary care: a qualitative study. *JMIR Mhealth Uhealth* 2015;3(4):e106 [FREE Full text] [doi: [10.2196/mhealth.4620](https://doi.org/10.2196/mhealth.4620)] [Medline: [26685288](#)]
37. Bailey JV, Webster R, Hunter R, Freemantle N, Rait G, Michie S, et al. The Men's Safer Sex (MenSS) trial: protocol for a pilot randomised controlled trial of an interactive digital intervention to increase condom use in men. *BMJ Open* 2015;5(2):e007552 [FREE Full text] [doi: [10.1136/bmjopen-2014-007552](https://doi.org/10.1136/bmjopen-2014-007552)] [Medline: [25687900](#)]
38. Muessig KE, Baltierra NB, Pike EC, LeGrand S, Hightow-Weidman LB. Achieving HIV risk reduction through HealthMpowerment.org, a user-driven eHealth intervention for young Black men who have sex with men and transgender women who have sex with men. *Digit Cult Educ* 2014;6(3):164-182 [FREE Full text] [Medline: [25593616](#)]

## Abbreviations

**HIV:** human immunodeficiency virus  
**IP:** Internet Protocol  
**LGBT:** lesbian, gay, bisexual, and transgender  
**MSM:** men who have sex with men  
**RCT:** randomized controlled trial  
**STI:** sexually transmitted infection

*Edited by G Eysenbach; submitted 22.07.16; peer-reviewed by V Minichiello, S Stahlman; comments to author 17.08.16; revised version received 21.08.16; accepted 21.08.16; published 12.09.16.*

*Please cite as:*

Mitchell J, Lee JY, Stephenson R

*How Best to Obtain Valid, Verifiable Data Online From Male Couples? Lessons Learned From an eHealth HIV Prevention Intervention for HIV-Negative Male Couples*

*JMIR Public Health Surveill* 2016;2(2):e152

URL: <http://publichealth.jmir.org/2016/2/e152/>

doi: [10.2196/publichealth.6392](https://doi.org/10.2196/publichealth.6392)

PMID: [27649587](https://pubmed.ncbi.nlm.nih.gov/27649587/)

©Jason Mitchell, Ji-Young Lee, Rob Stephenson. Originally published in *JMIR Public Health and Surveillance* (<http://publichealth.jmir.org>), 12.09.2016. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in *JMIR Public Health and Surveillance*, is properly cited. The complete bibliographic information, a link to the original publication on <http://publichealth.jmir.org>, as well as this copyright and license information must be included.

Original Paper

# Investigating Sociodemographic Factors and HIV Risk Behaviors Associated With Social Networking Among Adolescents in Soweto, South Africa: A Cross-Sectional Survey

Janan Janine Dietrich<sup>1</sup>, PhD; Fatima Laher<sup>1</sup>, MBBCh; Stefanie Hornschuh<sup>1</sup>, MSc; Busisiwe Nkala<sup>1</sup>, MHSc; Lucy Chimoyi<sup>1</sup>, MSc; Kennedy Otwombe<sup>1</sup>, MSc; Angela Kaida<sup>2</sup>, PhD; Glenda Elisabeth Gray<sup>1</sup>, MBBCh; Cari Miller<sup>2</sup>, PhD

<sup>1</sup>Perinatal HIV Research Unit (PHRU), Faculty of Health Sciences, University of the Witwatersrand, Johannesburg, South Africa

<sup>2</sup>Faculty of Health Sciences, Simon Fraser University, Vancouver, BC, Canada

**Corresponding Author:**

Janan Janine Dietrich, PhD  
Perinatal HIV Research Unit (PHRU)  
Faculty of Health Sciences  
University of the Witwatersrand  
Floor 12, Nurses Residence, Chris Hani Baragwanath Hospital  
Old Potch Road, Diepkloof, Soweto  
Johannesburg, 1864  
South Africa  
Phone: 27 119899831  
Fax: 27 86 218 0623  
Email: [dietrichj@phru.co.za](mailto:dietrichj@phru.co.za)

## Abstract

**Background:** Internet access via mobile phones and computers facilitates interaction and potential health communication among individuals through social networking. Many South African adolescents own mobile phones and can access social networks via apps.

**Objective:** We investigated sociodemographic factors and HIV risk behaviors of adolescent social networking users in Soweto, South Africa.

**Methods:** We conducted an interviewer-administered, cross-sectional survey of adolescents aged 14-19 years. Independent covariates of social networking were assessed by multivariate logistic regression analysis.

**Results:** Of 830 adolescents, 57% (475/830) were females and the median age was found to be 18 years (interquartile range 17-18). Social networking was used by 60% of adolescents (494/830); more than half, that is, 87% (396/494) accessed social networks through mobile phones and 56% (275/494) spent more than 4 hours per day using their mobile phones. Social networking was independently associated with mobile usage 2-4 hours (adjusted odds ratio [AOR]: 3.06, CI: 1.69-5.51) and more than 4 hours per day (AOR: 6.16, CI: 3.46-10.9) and one (AOR: 3.35, CI: 1.79-6.27) or more sexual partner(s) (AOR: 2.58, CI: 1.05-6.36).

**Conclusions:** Mobile phone-based social networking is prevalent among sexually active adolescents living in Soweto and may be used as an entry point for health promotion and initiation of low-cost adolescent health interventions.

(*JMIR Public Health Surveill* 2016;2(2):e154) doi:[10.2196/publichealth.4885](https://doi.org/10.2196/publichealth.4885)

**KEYWORDS**

mobile phone; adolescent health; HIV; health; social networking; mhealth; South Africa

## Introduction

Social networking offers the opportunity to offer low-cost, easily accessible information and interventions to reduce human immunodeficiency virus (HIV) risk among adolescents [1]. Given

the burden of HIV among young South Africans [2], novel uses of technology can be leveraged to disseminate information and interventions among adolescent social networking users [3]. Previous research conducted in developed settings show that HIV risk behaviors including depression, alcohol use, and risky

sexual behaviors are associated with social networking usage [4-6]. Despite a large body of research conducted in South Africa about HIV risk behaviors among adolescents [2,7,8], very few researchers have specifically investigated the association between HIV risk behaviors and mobile phone-based social networking.

Understanding mobile phone-based social networking in populations of interest may help in the relevant delivery of health interventions. In South Africa, a low-to-middle-income country, population census data reveals that 95% of households own mobile phones, compared with only 19% that own computers, desktops, or laptops [9]. In studies of South African adolescents and young adults, 72-78% own mobile phones whereas 91% have access to one, and up to 84% access the Internet via mobile phones [10,11]. The lower cost of Internet access through mobile phone broadband packages compared with computer-based Internet access is an important factor in the growth of mobile Internet in South Africa [11]. Taken together, these data suggest that health interventions using mobile phone technologies can reach the majority of the population in South Africa.

Internet access via mobile phones and computers facilitates interaction and potential health communication among individuals through social networking. Adolescents in particular have integrated into their lives the daily use of social media, including text messaging, blogging, videos sites, and social networking. Social networking platforms allow users to create online profiles to interact with each other [12,13]. In developed settings, social networking has been used to assess online social networking use and patterns [14], recruit participants for research participation [15,16], distribute questionnaires [17], assess problematic social networking use [18], promote health (including sexual health) [19,20], engage young people into treatment and care of HIV [21], and investigate the association between social networking use and sexual risk behaviors [4-6].

At the time of data collection for this study, the available social networking platforms in South Africa, included Mxit, Facebook, Whatsapp, and Blackberry messaging. Mxit, a South African innovation, is an instant messaging app that allows sending and receiving messages privately and via online chat rooms, while being able to access games, movie clips, or download music. Since Mxit was launched in 2003, it has registered at least 38 million South African users, which would equate to 73% of the total population in South Africa. Of the users, 19% are 15 to 17 years and 47% are 18 to 25 years old. The main activities on social networking platforms include sending messages, inviting others to events and online groups, and following the activities of other users [11]. These activities suggest that messages disseminated on social networking platforms have the potential to snowball, which is a useful feature for optimizing the reach of HIV prevention interventions among young people. When one considers the penetration of social networking platforms in African settings, the relevance of leveraging them for health promotion interventions is evident [22].

A few studies, mainly from the developed world, have examined user characteristics of social networking platforms. The factors that have been investigated include demographics, sexual

behavior, and substance use [23-25]. For example, a study among 1434 10<sup>th</sup> grade students from 5 high schools in Southern California showed that higher levels of alcohol use were associated with sending and receiving friendship invitations, becoming friends with users that consume alcohol and also increasing alcohol use if their friends drank more. In addition, results revealed that exposure to risky online pictures posted by friends increased smoking behaviors amongst study participants [23]. Behavioral risk factors such as sexual partnerships and alcohol use have also been examined with respect to social networking platform use. Clayton et al (2013) indicated that alcohol use among undergraduate students from the United States was a predictor for emotional attachment to Facebook [26].

Developers in South Africa have attempted to harness mobile phone technology for health message delivery. For example, in 2008 LoveLife launched the first mobile social networking app in South Africa, called MyMsta, which was dedicated to youth empowerment and HIV prevention [27]. However, there is little published evidence on the effectiveness and cultural acceptability of these types of health interventions in developing world settings like South Africa, particularly among adolescents [22]. Adolescents in developing settings face the enormous challenge of the intersection between infectious and noninfectious disease epidemics in these countries. Understanding the characteristics of adolescents accessing social networking platforms in developing countries may lend insight into the development of appropriate and targeted health interventions. Few studies have specifically focused on the use of social networking among adolescents in South Africa and even fewer have published results of these studies. This study investigated demographic, socioeconomic, sexual behavior, alcohol use, depression, and mobile phone use associated with social networking amongst adolescents in Soweto, South Africa.

## Methods

### Study Design

This analysis forms part of the *Botsha Bophelo Adolescent Health Study (BBAHS)*, an interviewer-administered, cross-sectional survey of 830 adolescents aged 14-19 years living in Soweto.

### Setting

The study was conducted at the Perinatal HIV Research Unit (PHRU) and at Kganya Motsha Adolescent Centre (KMAC) in Kliptown. Kliptown is the oldest district in Soweto, and it comprises a mixture of purpose-built housing and informal housing known as shacks [28]. Soweto is a large peri-urban township of about 1.3 million official inhabitants [29]. KMAC was founded in 2008 as an adolescent-friendly HIV management centre serving the HIV voluntary counseling and testing (VCT) and sexual and reproductive health needs of adolescents aged 14-19 years [30].

### Participants

A targeted, stratified sampling approach was used based on geographic distribution across all areas of Soweto, as well as

the sex and age of adolescents. Seven of the areas were informal areas (locally known as informal settlements), where adolescents lived in housing structures composed of metal structures (locally known as shacks), with shared access to water, toilets, and limited or no access to electricity. Eligibility criteria included age 14-19 years and living in Soweto. Participants were recruited from various locations within each area with the aim of ensuring representation from all areas of Soweto and recruiting hard-to-reach adolescents who may not have accessed health services. Research interviewers strategically targeted adolescents around malls, schools, and neighbourhood hangouts. Interviewers provided potential participants with recruitment flyers with contact information where interviewers could be reached. Interviewers were available on a drop-in basis at KMAC and the PHRU or by appointment. To accommodate adolescent schedules, interview times were available after school and on the weekends. Some communities were more challenging to recruit from than others, particularly informal settlements where the visibility of adolescents was reduced due to the lack of community infrastructure for young people to congregate (ie, schools and malls). Participants learned about the study from research staff and were invited to follow-up if they were interested in participating. Participants were also invited to share the information about the study with their friends. Therefore, we have a convenience sample based on targeted recruitment strategies stratified by geographic location, age, and sex characteristics.

### Data Collection

Surveys were conducted between June 2010 and June 2012. Participants received a face-to-face interviewer-administered survey assessing demographics, mobile phone access and use, sexual risk behavior, and alcohol use. Trained, multilingual interviewers administered the survey in English or IsiZulu (the most commonly used local languages in Soweto) via an online Internet platform, Survey Monkey, using iPads or computers [16,31]. Each survey took about 60 minutes.

### Ethical Considerations

The ethics committees of the University of the Witwatersrand, Johannesburg, South Africa and Simon Fraser University, Vancouver, Canada approved study procedures. Participants received 50 ZAR (~ \$7) for travel reimbursement. Written informed consent was obtained for all participants. In addition, participants younger than 18 years required written parental consent together with their own written assent.

### Measures

The primary outcome variable of interest in this analysis was social networking, defined as the use of social networking platforms such as Facebook, Twitter, and Mxit accessed through mobile phones. Social networking was measured by responses (yes vs no). Explanatory variables included demographics

(including: gender, age, primary language spoken at home, education level [primary vs post primary]), socioeconomic (type of housing, source of spending money), sexual behavior (sexual orientation, sexually active, number of sexual partnerships, same sex partnerships, age at coital debut, condom use, intergenerational sex), alcohol and drug use in the last 6 months, depression, mobile phone use (eg, access to a mobile phone, access to the Internet, and source of Internet [computer vs mobile phone]), sexual partnerships ([none, one, and more than one]), age at coital debut [less than 15 years, 15-16 years, and 17 or more years], intergenerational sex and condom use), and relational (eg, parental and adult presence in household) variables.

### Statistical Analysis

Participant characteristics of adolescents who accessed social networking sites are presented in frequencies and compared through bivariate analysis using chi-square and Fisher's exact test. Variables with a significant association ( $P < .05$ ) were considered for entry into the multivariate logistic regression model. Univariate and multivariate logistic regression were performed to determine crude and adjusted odds ratio (AOR), 95% confidence interval (CI), and their corresponding  $P$  values. Model fit was assessed using the Hosmer-Lemeshow goodness-of-fit statistic where the model was determined a good fit if the  $P$  value was not significant [32]. All statistical tests were two-sided and were conducted using STATA version 12.

## Results

### Demographic Characteristics

In total, 956 interviews were completed between 2010 and 2012. Following extensive data cleaning, a final sample of 830 adolescent participants was reached. Due to the nature of our recruitment strategy, we were not able to determine absolute response or refusal rates. The median age for the sample ( $n=830$ ) was 18 years (IQR 17-18) and 57% ( $n=475$ ) were female. Of all participants, 60% ( $n=494$ ) used social networking platforms. The most commonly used platforms were Mxit (46%) and Facebook (34%).

Table 1 shows the comparison of characteristics of social networking users ( $n=494$ ) to nonsocial networking users ( $n=336$ ). In bivariate analyses, social networking was significantly associated with sex ( $P=.041$ ), age group ( $P=.002$ ), type of housing ( $P<.001$ ), educational level ( $P<.001$ ), access to mobile phone ( $P<.001$ ), access to Internet ( $P<.001$ ), and source of Internet ( $P<.001$ ). Other variables associated with social networking were alcohol use in the past 6 months ( $P<.001$ ), daily mobile phone usage ( $P<.001$ ), presence of an adult in the household ( $P=.001$ ), sexually active ( $P=.032$ ), sexual partnerships ( $P=.011$ ), and age at coital debut ( $P=.012$ ).

**Table 1.** Summary profile of adolescents by social networking use (N=830), Soweto, South Africa.

Variable	N (830)	Social networking nonusers, n=336 (40.0%)	Social networking users, n=494 (60.0%)	P value
<b>Demographic</b>				
<b>Sex</b>				
Male	355	158 (47.0)	197 (39.9)	.041
Female	475	178 (53.0)	297 (60.1)	
<b>Age group</b>				
Less than 15 years	179	90 (27.5)	89 (18.0)	.002
16-17 years	233	95 (29.1)	138 (28.0)	
18-19 years	408	142 (43.4)	266 (54.0)	
<b>Language</b>				
IsiZulu	446	186 (56.4)	260 (53.2)	.138
IsiXhosa	75	36 (10.9)	39 (7.9)	
Sesotho	120	39 (11.8)	81 (16.6)	
Others	178	69 (20.9)	109 (22.3)	
<b>Education level</b>				
Primary	198	101 (30.6)	97 (19.6)	<.001
Post Primary	626	229 (69.4)	397 (80.4)	
<b>Socioeconomic</b>				
<b>Type of housing</b>				
Formal	700	261 (78.4)	439 (89.8)	<.001
Informal	122	72 (21.6)	50 (10.2)	
<b>Source of spending money</b>				
Parents	647	254 (78.4)	392 (82.4)	.163
Other (Employment, relatives, boy or girlfriend)	154	70 (21.6)	84 (17.6)	
<b>Sexual behavior</b>				
<b>Sexual orientation</b>				
Heterosexual	686	267 (86.4)	419 (87.3)	.719
Lesbian or gay or bisexual	103	42 (13.6)	61 (12.7)	
<b>Sexually active</b>				
No	369	164 (48.8)	205 (41.6)	.032
Yes	460	172 (51.2)	288 (58.4)	
<b>Sexual partnerships<sup>a,b</sup></b>				
None	70	36 (15.8)	34 (8.4)	.011
One	366	130 (57.0)	236 (58.3)	
More than one	197	62 (27.2)	135 (33.3)	
<b>Same sex partnerships<sup>a,b</sup></b>				
No	269	90 (90.9)	179 (95.7)	.101
Yes	17	9 (9.1)	8 (4.3)	
<b>Age at coital debut<sup>a,b</sup></b>				
Less than 15 years	95	36 (25.2)	59 (23.1)	.012
15-16 years	164	70 (49.0)	94 (36.9)	



Variable	N (830)	Social networking nonusers, n=336 (40.0%)	Social networking users, n=494 (60.0%)	P value
17 years or more	139	37 (25.9)	102 (40.0)	
<b>Condom use<sup>a,b</sup></b>				
Never	15	8 (8.9)	7 (3.9)	.161
Sometimes	68	25 (27.8)	43 (24.0)	
Always	186	57 (63.3)	129 (72.1)	
<b>Intergenerational sex<sup>a,b</sup></b>				
No	523	190 (87.6)	333 (84.3)	.275
Yes	89	27 (12.4)	62 (15.7)	
<b>Alcohol and drug use in the last six months</b>				
<b>Alcohol use in the last six months</b>				
No	294	150 (45.3)	144 (29.5)	<.001
Yes	525	181 (54.7)	344 (70.5)	
<b>Drug use in the last six months</b>				
No	714	331 (98.5)	483 (98.0)	.567
Yes	15	5 (1.5)	10 (2.0)	
<b>Depression</b>				
No	542	227 (67.6)	315 (63.9)	.276
Yes	287	109 (32.4)	178 (36.1)	
<b>Mobile phone use</b>				
<b>Access to Internet</b>				
No	253	214 (72.1)	39 (7.9)	<.001
Yes	535	83 (28.0)	452 (92.1)	
<b>Source of Internet</b>				
Computer	87	28 (32.6)	59 (13.0)	<.001
Mobile phone	454	58 (67.4)	396 (87.0)	
<b>Access to mobile phones</b>				
No	148	115 (38.1)	33 (6.7)	<.001
Yes	648	187 (61.9)	461 (93.3)	
<b>Daily mobile phone usage</b>				
0-1 hour	226	148 (49.8)	78 (15.9)	<.001
2-4 hours	217	80 (26.9)	137 (28.0)	
>4 hours	344	69 (23.3)	275 (56.1)	
<b>Relational or parental</b>				
<b>Parental presence</b>				
No parents	49	23 (6.8)	26 (5.3)	.085
One parent	296	132 (39.3)	164 (33.2)	
Both parents	485	181 (53.9)	304 (61.5)	
<b>Presence of an adult in household</b>				
No	125	67 (20.3)	58 (12.0)	.001

Variable	N (830)	Social networking nonusers, n=336 (40.0%)	Social networking users, n=494 (60.0%)	P value
Yes	696	263 (79.7)	427 (88.0)	

<sup>a</sup>Considers only those who responded to the item.

<sup>b</sup>Column proportions do not add up to 100% due to missing values in some variables.

### Characteristics Associated With Social Networking Use

The unadjusted and adjusted predictors of social networking use are presented in Table 2. In the adjusted logistic regression

model, adolescents who used mobile phones for 2-4 hours (AOR: 2.89, CI: 1.80-4.65) or more than 5 hours (AOR: 5.99, CI: 3.79-9.48) daily and those with one (AOR: 1.75, CI: 1.04-2.94) or more than one (AOR: 2.64, CI: 1.40-4.96) sexual partner had a higher odds of social networking.

**Table 2.** Unadjusted and adjusted associations between selected variables and social networking among adolescents, Soweto (N=830).

Factor	Univariate			Multivariate		
	OR	95% CI	P value	AOR	95% CI	P value
<b>Sex</b>						
Male	1.00 (Ref <sup>a</sup> )			-	-	-
Female	1.34	1.01-1.77	.04	-	-	-
<b>Age group</b>						
Less than 15 years	1.00 (Ref)					
16-17 years	1.47	0.99-2.16	.06	-	-	-
18-19 years	1.89	1.33-2.71	<.001	-	-	-
<b>Source of spending money</b>						
Parents	1.00 (Ref)			-	-	-
Others	0.77	0.55-1.11	.164	-	-	-
<b>Education level</b>						
Primary	1.00 (Ref)					
Post primary	1.85	1.31-2.49	<.001	-	-	-
<b>Type of housing</b>						
Formal	1.00 (Ref)					
Informal	0.41	0.23-0.61	<.001	-	-	-
<b>Presence of an adult in the household</b>						
No	1.00 (Ref)					
Yes	1.88	1.28-2.75	<.001	-	-	-
<b>Daily mobile phone usage</b>						
0-1 hour	1.00 (Ref)			1.00 (Ref)		
2-4 hours	3.27	2.22-4.83	<.001	2.89	1.80-4.65	<.001
>4 hours	7.56	5.17-11.06	<.001	5.99	3.79-9.48	<.001
<b>Depressed</b>						
No	1.00 (Ref)					
Yes	1.18	0.88-1.58	.276	-	-	-
<b>Sexual partnerships</b>						
None	1.00 (Ref)			1 (Ref)		
One	1.92	1.15-3.22	0.01	1.75	1.04-2.94	.035
More than One	2.31	1.32-4.02	<.001	2.64	1.40-4.96	.003

<sup>a</sup>Ref: Reference group.

The unadjusted and adjusted predictors of social networking use by sex are presented in Table 3. In the adjusted logistic regression by sex, males whose source of spending money was not provided by parents (AOR: 0.58, CI: 0.34-0.99) had lower odds of social networking. Those who used mobile phones for 2-4 hours (AOR: 3.06, CI: 1.69-5.51) or more than 4 hours daily (AOR: 6.16, CI: 3.46-10.9) had a higher odds of social

networking. Adolescent males with one (AOR: 3.35, CI: 1.79-6.27) or more than one (AOR: 2.58, CI: 1.05-6.36) sexual partner had a higher odds of networking. Among females, a higher odds of social networking was associated with spending 2-4 hours (AOR: 3.53, CI: 2.08-5.97) or more than 5 hours (AOR: 6.49, CI: 3.92-10.73) daily on mobile phones and having more than one sexual partner (AOR: 1.99, CI: 1.08-3.67).

**Table 3.** Unadjusted and adjusted associations between selected variables and social networking among adolescents by sex, Soweto.

Factor	Males				Females			
	Univariate		Multivariate		Univariate		Multivariate	
	OR (95% CI)	P value	AOR (95% CI)	P value	OR (95% CI)	P value	AOR (95% CI)	P value
<b>Age group</b>								
Less than 15 years	1.00 (Ref <sup>a</sup> )				1.00 (Ref)			
16-17 years	1.24 (0.69-2.23)	.463	-	-	1.61 (0.94-2.75)	.080	-	-
18-19 years	1.82 (1.08-3.09)	.026	-	-	1.90 (1.16-3.09)	.010	-	-
<b>Type of Housing</b>								
Formal	1.00 (Ref)				1.00 (Ref)			
Informal	0.50 (0.28-0.92)	.025	-	-	0.36 (0.21-0.60)	<.001	-	-
<b>Education level</b>								
Primary	1.00 (Ref)				1.00 (Ref)			
Post Primary	1.56 (0.98-2.75)	.058	-	-	1.94 (1.23-3.07)	.004	-	-
<b>Source of spending money</b>								
Parents	1.00 (Ref)		1.00 (Ref)		1.00 (Ref)		-	-
Others	0.60 (0.36-1.00)	.052	0.58 (0.34-0.99)	.045	1.02 (0.62-1.68)	.939	-	-
<b>Daily mobile phone usage</b>								
0-1 hour	1.00 (Ref)				1.00 (Ref)		1.00 (Ref)	
2-4 hours	2.84 (1.60-5.06)	<.001	3.06 (1.69-5.51)	<.001	3.45 (2.03-5.86)	<.001	3.53 (2.08-5.97)	<.001
>4 hours	6.24 (3.56-10.87)	<.001	6.16 (3.46-10.9)	<.001	8.74 (5.17-14.75)	<.001	6.49 (3.92-10.73)	<.001
<b>Sexual Partnerships</b>								
None	1.00 (Ref)				1.00 (Ref)		1.00 (Ref)	
One	1.61 (0.78-3.34)	.196	3.35 (1.79-6.27)	.036	2.28 (1.29-4.06)	.005	1.47 (0.91-2.38)	.12
More than One	2.28 (1.08-4.84)	.031	2.58 (1.05-6.36)	.0002	4.94 (1.98-12.37)	.001	1.99 (1.08-3.67)	.027
<b>Presence of an adult in the household</b>								
No	1.00 (Ref)				1.00 (Ref)			
Yes	1.14 (0.64-2.03)	.653	-	-	2.77 (1.63-4.69)	<.001	-	-
<b>Depressed</b>								
No	1.00 (Ref)				1.00 (Ref)			
Yes	1.19 (0.76-1.87)	.450	-	-	1.13 (0.77-1.68)	.511	-	-

<sup>a</sup>Ref: Reference group.

## Discussion

### Principal Findings

This is one of the first studies to report on social networking use via mobile phones among adolescents in South Africa. The data show that social networking use among a sample of

adolescents in Soweto was associated with their source of spending money from either parents or others, longer duration of daily mobile phone use, and having more than one sexual partnership. In an analysis by sex, social networking was associated in both sexes with longer daily use of mobile phones, being sexually active and having multiple sexual partners.

Interestingly, for females, there was a relationship between social networking use and alcohol use. For males, social networking use was associated with longer daily use of mobile phones and having more than one sexual partner.

Despite Soweto being a typically lower socioeconomic community in South Africa, social network access was prevalent among 60% of adolescents, indicating the opportunity to introduce HIV risk reduction interventions via social networking platforms to young people in this setting. Mobile phones were the main point of social networking access for the majority of the adolescents. In developed settings like the United States, social networking has been shown to be a feasible tool for health research among adolescents [33]. Studies (Yonker et al, 2015) have also shown that using social networking among adolescents and young adults concerning their health choices has proven to be an essential medium of communication [34]. Magidson et al (2015) demonstrated the effective use of social networking to engage HIV positive adult men who have sex with men in care [35]. In sub-Saharan Africa, published literature on social networking among adolescent samples is limited.

In our study, we found that access to social networking was lower among adolescents with lower socioeconomic markers including less educational attainment, living in informal housing arrangements and no adult presence in the household. We also found that those adolescents accessing social networking were more likely to be older, sexually active, and using alcohol. Taken together, these findings indicate the excellent potential to reach adolescents in Soweto with HIV prevention interventions via social networking, particularly those with higher risk sexual behaviors including being sexually active, having more than one sexual partner, and those who use alcohol. However, it may be harder to reach adolescents facing lower socioeconomic hardships via social networking.

Among participants under age 18 years, almost a third reported alcohol use, although we did not measure the extent of use. Findings from the second South African youth risk behavior survey (2008) showed that half of school-going adolescents aged 13–19 years had ever used alcohol and a third having consumed alcohol in the past month. Two-thirds of social networking users reported alcohol use in the past 6 months. Of those who reported social networking, almost half were underage. Furthermore, half of female adolescents who consumed alcohol in the past 6 months were more likely to use social networking platforms [36].

The link between alcohol advertising and use among adolescents has been documented before; however, research conducted in the United States has shown conflicting evidence about the effect of media campaigns to reduce problematic alcohol use. There is evidence that links alcohol advertising with alcohol

initiation [37]. Furthermore, there is evidence from the United States that exposure to friends' accounts of alcohol use on social networking platforms could have an indirect effect on willingness to consume alcohol [38]. By extension it may be reasonable to assume that if adolescents were exposed to normative antialcohol interventions on social networking platforms, they may in turn become less likely to initiate or continue alcohol use. In this way, social networking platforms could allow targeting of specific populations for novel antialcohol interventions. Male adolescents are more likely to use alcohol at an early age but there has been an increase in the number of younger female adolescents consuming alcohol [39,40], which is a concerning trend because females are physiologically more at risk from the effects of alcohol use than males [41,42,43]. Young females are an emerging target market for alcohol producers [40]. Adolescent alcohol use has been described as a factor in female sexual assault and engaging in risky sexual behaviors [44,45]. Among social networking users, almost a third reported multiple sexual partnerships. Adolescents with multiple sexual partners are a higher risk group for sexually transmitted infections [2]. However, multiple and concurrent sexual partnering is a risk factor for HIV only if condom use is inconsistent and if a partner is infected with HIV [46]. In South Africa, multiple and concurrent sexual partnerships has been associated with HIV infection among young people [2,47-50]. Of sexually active young people aged 15-24 years in South Africa, 22% had more than one sexual partner in the past 12 months [47]. Social networking applications can potentially be used to promote positive health messages through adolescent social networks.

### Limitations

This study did not document the type of content that adolescents were accessing on the social networking platforms; therefore, we could not determine whether adolescents demonstrated health-seeking behaviors on these platforms. Social desirability bias may have affected variables, possibly underestimating self-reported risk behaviors. Social networking based health applications may not be broadly applicable but would be effective for subsets of the adolescent population who are able to obtain data bundles to access social networking sites. There were limitations to the survey items because we did not assess the level of alcohol use and concurrency of sexual partnerships.

### Conclusion

Our study suggests evidence about mobile phone-based social networking use among adolescents in Soweto, South Africa, and beyond. We recommend that future research addresses social networking platforms as a means of disseminating health interventions including positive alcohol and sexual behavior messaging among adolescent populations in South Africa.

### Acknowledgments

The investigators of the Botsha Bophelo study wish to thank the adolescent participants for generously sharing their time and insights with us. We also wish to thank the PHRU and Kganya Motsha Adolescent Centre for their support in this project. The study was funded by the Canadian Institutes of Health Research (CIHR), Human Development, Child and Youth Health: 230513, and the South African AIDS Vaccine Initiative (SAVI). Initial seed funding was provided by Simon Fraser University (SFU)

through a President Research Award to Dr. Cari Miller. Dr. Janan Dietrich holds Thuthuka post PhD funding from the South African National Research Foundation (NRF). We would like to thank the research staff at PHRU who contributed to recruitment and data collection efforts. Any opinion, finding and conclusion or recommendation expressed in this material is that of the authors and the SA NRF does not accept any liability in this regard.

## Conflicts of Interest

None declared.

## References

1. Taggart T, Grewe ME, Conserve DF, Gliwa C, Isler MR. Social Media and HIV: A Systematic Review of Uses of Social Media in HIV Communication. *J Med Internet Res* 2015;17(11):e248 [FREE Full text] [doi: [10.2196/jmir.4387](https://doi.org/10.2196/jmir.4387)] [Medline: [26525289](https://pubmed.ncbi.nlm.nih.gov/26525289/)]
2. Shisana O, Rehle T, Simbayi LC, Zuma K, Jooste S, Zungu N, et al. South African National HIV Prevalence, Incidence and Behaviour Survey, 2012. Cape Town, South Africa: HSRC Press; 2014.
3. Baelden D, Van Audenhove L, Vergnani T. Using new technologies for stimulating interpersonal communication on HIV and AIDS. *Telematics and Informatics* 2012 May;29(2):166-176. [doi: [10.1016/j.tele.2011.05.002](https://doi.org/10.1016/j.tele.2011.05.002)]
4. Hughes D, Rowe M, Batey M, Lee A. A tale of two sites: Twitter vs. Facebook and the personality predictors of social media usage. *Computers in Human Behavior* 2012 Mar;28(2):561-569. [doi: [10.1016/j.chb.2011.11.001](https://doi.org/10.1016/j.chb.2011.11.001)]
5. Mahajan A, Sayles J, Patel V, Remien R, Sawires SR, Ortiz D, et al. Stigma in the HIV/AIDS epidemic: a review of the literature and recommendations for the way forward. *AIDS* 2008 Aug;22(Suppl 2):S67-S79 [FREE Full text] [doi: [10.1097/01.aids.0000327438.13291.62](https://doi.org/10.1097/01.aids.0000327438.13291.62)] [Medline: [18641472](https://pubmed.ncbi.nlm.nih.gov/18641472/)]
6. Blackstock OJ, Houghton LJ, Garner RY, Horvath KJ, Norwood C, Cunningham CO. General and health-related Internet use among an urban, community-based sample of HIV-positive women: implications for intervention development. *AIDS Care* 2015;27(4):536-544 [FREE Full text] [doi: [10.1080/09540121.2014.980215](https://doi.org/10.1080/09540121.2014.980215)] [Medline: [25411825](https://pubmed.ncbi.nlm.nih.gov/25411825/)]
7. Dietrich J, Khunwane M, Laher F, de Bruyn BG, Sikkema K, Gray G. "Group sex" parties and other risk patterns: A qualitative study about the perceptions of sexual behaviors and attitudes of adolescents in Soweto, South Africa. *Vulnerable Child Youth Stud* 2011 Sep 1;6(3):244-254 [FREE Full text] [doi: [10.1080/17450128.2011.597796](https://doi.org/10.1080/17450128.2011.597796)] [Medline: [22844344](https://pubmed.ncbi.nlm.nih.gov/22844344/)]
8. Dietrich J, Sikkema K, Otjombe KN, Sanchez A, Nkala B, de BG, Van Der Watt Martin, et al. Multiple levels of influence in predicting sexual activity and condom use among adolescents in Soweto, Johannesburg, South Africa. *J HIV AIDS Soc Serv* 2013 Jul 1;12(3-4):404-423 [FREE Full text] [doi: [10.1080/15381501.2013.819312](https://doi.org/10.1080/15381501.2013.819312)] [Medline: [24532992](https://pubmed.ncbi.nlm.nih.gov/24532992/)]
9. Statistics South Africa. General household survey. In: Statistical Release. Pretoria, South Africa: Statistics South Africa; 2012.
10. Dietrich JJ, Coetzee J, Otjombe K, Mdanda S, Nkala B, Makongoza M, et al. Adolescent-friendly technologies as potential adjuncts for health promotion. *Health Education* 2014 Jun 02;114(4):304-318. [doi: [10.1108/HE-08-2013-0039](https://doi.org/10.1108/HE-08-2013-0039)]
11. UNICEF New York, Division of Communication, Social and Civic Media Section, Beger G, Sinha A. South African mobile generation: Study on South African young people on mobiles. In: UNICEF. New York: UNICEF; May 29, 2012.
12. Boyd DM, Ellison NB. Social Network Sites: Definition, History, and Scholarship. *Journal of Computer-Mediated Communication* 2007 Oct;13(1):210-230 [FREE Full text] [doi: [10.1111/j.1083-6101.2007.00393.x](https://doi.org/10.1111/j.1083-6101.2007.00393.x)]
13. Wyche SP, Schoenebeck SY, Forte A. "Facebook is a Luxury": An exploratory study of social media use in rural Kenya. In: *Into the Petri Dish: Culture and Collaboration*. San Antonio, TX: CSCW; Feb 2013:33-44.
14. Tsitsika AK, Tzavela EC, Janikian M, Ólafsson K, Iordache A, Schoenmakers TM, et al. Online social networking in adolescence: patterns of use in six European countries and links with psychosocial functioning. *J Adolesc Health* 2014 Jul;55(1):141-147. [doi: [10.1016/j.jadohealth.2013.11.010](https://doi.org/10.1016/j.jadohealth.2013.11.010)] [Medline: [24618179](https://pubmed.ncbi.nlm.nih.gov/24618179/)]
15. Amon K, Campbell A, Hawke C, Steinbeck K. Facebook as a Recruitment Tool for Adolescent Health Research: A Systematic Review. *Academic Pediatrics* 2014 Sep;14(5):439-447.e4. [doi: [10.1016/j.acap.2014.05.049](https://doi.org/10.1016/j.acap.2014.05.049)]
16. Harris ML, Loxton D, Wigginton B, Lucke JC. Recruiting online: lessons from a longitudinal survey of contraception and pregnancy intentions of young Australian women. *Am J Epidemiol* 2015 May 15;181(10):737-746 [FREE Full text] [doi: [10.1093/aje/kwv006](https://doi.org/10.1093/aje/kwv006)] [Medline: [25883155](https://pubmed.ncbi.nlm.nih.gov/25883155/)]
17. Ljepava N, Orr R, Locke S, Ross C. Personality and social characteristics of Facebook non-users and frequent users. *Computers in Human Behavior* 2013 Jul;29(4):1602-1607. [doi: [10.1016/j.chb.2013.01.026](https://doi.org/10.1016/j.chb.2013.01.026)]
18. Jelenchick L, Hawk S, Moreno M. Problematic internet use and social networking site use among Dutch adolescents. *Int J Adolesc Med Health* 2016 Feb;28(1):119-121. [doi: [10.1515/ijamh-2014-0068](https://doi.org/10.1515/ijamh-2014-0068)] [Medline: [25720115](https://pubmed.ncbi.nlm.nih.gov/25720115/)]
19. Syred J, Naidoo C, Woodhall SC, Baraitser P. Would you tell everyone this? Facebook conversations as health promotion interventions. *J Med Internet Res* 2014;16(4):e108 [FREE Full text] [doi: [10.2196/jmir.3231](https://doi.org/10.2196/jmir.3231)] [Medline: [24727742](https://pubmed.ncbi.nlm.nih.gov/24727742/)]
20. Bull S, Levine D, Black S, Schmiege S, Santelli J. Social Media-Delivered Sexual Health Intervention. *American Journal of Preventive Medicine* 2012 Nov;43(5):467-474. [doi: [10.1016/j.amepre.2012.07.022](https://doi.org/10.1016/j.amepre.2012.07.022)]
21. Hailey Jamal H N, Arscott J. Using technology to effectively engage adolescents and young adults into care: STAR TRACK Adherence Program. *J Assoc Nurses AIDS Care* 2013;24(6):582-586. [doi: [10.1016/j.jana.2013.03.001](https://doi.org/10.1016/j.jana.2013.03.001)] [Medline: [23809658](https://pubmed.ncbi.nlm.nih.gov/23809658/)]



22. 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](https://doi.org/10.1371/journal.pmed.1001382)] [Medline: [23424286](https://pubmed.ncbi.nlm.nih.gov/23424286/)]
23. Huang GC, Soto D, Fujimoto K, Valente TW. The interplay of friendship networks and social networking sites: longitudinal analysis of selection and influence effects on adolescent smoking and alcohol use. *Am J Public Health* 2014 Aug;104(8):e51-e59. [doi: [10.2105/AJPH.2014.302038](https://doi.org/10.2105/AJPH.2014.302038)] [Medline: [24922126](https://pubmed.ncbi.nlm.nih.gov/24922126/)]
24. Vo K, Smollin C. Online social networking and US poison control centers: Facebook as a means of information distribution. *Clin Toxicol (Phila)* 2015 Jun;53(5):466-469. [doi: [10.3109/15563650.2015.1014906](https://doi.org/10.3109/15563650.2015.1014906)] [Medline: [25761679](https://pubmed.ncbi.nlm.nih.gov/25761679/)]
25. Brown J, Michie S, Raupach T, West R. Prevalence and characteristics of smokers interested in internet-based smoking cessation interventions: cross-sectional findings from a national household survey. *J Med Internet Res* 2013;15(3):e50 [FREE Full text] [doi: [10.2196/jmir.2342](https://doi.org/10.2196/jmir.2342)] [Medline: [23506944](https://pubmed.ncbi.nlm.nih.gov/23506944/)]
26. Clayton RB, Osborne RE, Miller BK, Oberle CD. Loneliness, anxiousness, and substance use as predictors of Facebook use. *Computers in Human Behavior* 2013 May;29(3):687-693. [doi: [10.1016/j.chb.2012.12.002](https://doi.org/10.1016/j.chb.2012.12.002)]
27. Dube A, Burnett S, Mymsta. Sandton, South Africa: LoveLife; 2010. Using mobile social networks to build social capital and reduce risk tolerance among young people in South Africa URL: <http://www.mymsta.mobi/> [accessed 2016-09-08] [WebCite Cache ID 6kN2KLgxw]
28. Estrera AL, Miller CC, Guajardo-Salinas G, Coogan S, Charlton-Ouw K, Safi HJ, et al. Update on blunt thoracic aortic injury: fifteen-year single-institution experience. *J Thorac Cardiovasc Surg* 2013 Mar;145(3 Suppl):S154-S158. [doi: [10.1016/j.jtcvs.2012.11.074](https://doi.org/10.1016/j.jtcvs.2012.11.074)] [Medline: [23260456](https://pubmed.ncbi.nlm.nih.gov/23260456/)]
29. City of Johannesburg Metropolitan Municipality. Joburg. Johannesburg, South Africa: City of Johannesburg Metropolitan Municipality; 2006. The remaking of Soweto End of Term Report 2006 – 2011 URL: <http://www.joburg.org.za/> [accessed 2016-09-08] [WebCite Cache ID 6ZhH7t8Vi]
30. Nkala B, Khunwane M, Dietrich J, Otwombe K, Sekoane I, Sonqishe B, et al. Kganya Motsha Adolescent Centre: a model for adolescent friendly HIV management and reproductive health for adolescents in Soweto, South Africa. *AIDS Care* 2015;27(6):697-702. [doi: [10.1080/09540121.2014.993352](https://doi.org/10.1080/09540121.2014.993352)] [Medline: [25599113](https://pubmed.ncbi.nlm.nih.gov/25599113/)]
31. SurveyMonkey Inc.. Survey Monkey. Palo Alto, CA URL: <http://www.surveymonkey.com/> [accessed 2016-09-08] [WebCite Cache ID 6ZhHKof7z]
32. Hosmer D, Lemeshow S. Applied logistic regression. New York: Wiley; 2000.
33. Brockman LN, Christakis DA, Moreno MA. Friending Adolescents on Social Networking Websites: A Feasible Research Tool. *J Interact Sci* 2014;2(1) [FREE Full text] [doi: [10.1186/2194-0827-2-1](https://doi.org/10.1186/2194-0827-2-1)] [Medline: [25485226](https://pubmed.ncbi.nlm.nih.gov/25485226/)]
34. Yonker LM, Zan S, Scirica CV, Jethwani K, Kinane TB. "Friending" teens: systematic review of social media in adolescent and young adult health care. *J Med Internet Res* 2015;17(1):e4 [FREE Full text] [Medline: [25560751](https://pubmed.ncbi.nlm.nih.gov/25560751/)]
35. Magidson JF, Biello KB, Safren SA, Rosenberger JG, Novak DS, Mayer KH, et al. Engagement in HIV care and sexual transmission risk behavior among men who have sex with men using online social/sexual networking in Latin America. *AIDS Care* 2015 Aug;27(8):1055-1062. [doi: [10.1080/09540121.2015.1017796](https://doi.org/10.1080/09540121.2015.1017796)] [Medline: [25738655](https://pubmed.ncbi.nlm.nih.gov/25738655/)]
36. Reddy S, James S, Sewpaul R, Koopman F, Funani N, Sifunds S, et al. Umthente Uhlaba Usamila - The South African Youth Risk Behaviour Survey. In: Umthente Uhlaba Usamila. Cape Town, South Africa: South African Medical Research Council; 2010.
37. Grenard JL, Dent CW, Stacy AW. Exposure to alcohol advertisements and teenage alcohol-related problems. *Pediatrics* 2013 Feb;131(2):e369-e379 [FREE Full text] [doi: [10.1542/peds.2012-1480](https://doi.org/10.1542/peds.2012-1480)] [Medline: [23359585](https://pubmed.ncbi.nlm.nih.gov/23359585/)]
38. Litt DM, Stock ML. Adolescent alcohol-related risk cognitions: the roles of social norms and social networking sites. *Psychol Addict Behav* 2011 Dec;25(4):708-713. [doi: [10.1037/a0024226](https://doi.org/10.1037/a0024226)] [Medline: [21644803](https://pubmed.ncbi.nlm.nih.gov/21644803/)]
39. Roche AM, Deehan A. Women's alcohol consumption: emerging patterns, problems and public health implications. *Drug Alcohol Rev* 2002 Jun;21(2):169-178. [doi: [10.1080/09595230220139073](https://doi.org/10.1080/09595230220139073)] [Medline: [12188996](https://pubmed.ncbi.nlm.nih.gov/12188996/)]
40. Flegel K. Big alcohol catches up with adolescent girls. *CMAJ* 2013 Jul 9;185(10):859 [FREE Full text] [doi: [10.1503/cmaj.130766](https://doi.org/10.1503/cmaj.130766)] [Medline: [23754102](https://pubmed.ncbi.nlm.nih.gov/23754102/)]
41. Agrawal A, Grant JD, Waldron M, Duncan AE, Scherrer JF, Lynskey MT, Madden Pamela A F, et al. Risk for initiation of substance use as a function of age of onset of cigarette, alcohol and cannabis use: findings in a Midwestern female twin cohort. *Prev Med* 2006 Aug;43(2):125-128. [doi: [10.1016/j.ypmed.2006.03.022](https://doi.org/10.1016/j.ypmed.2006.03.022)] [Medline: [16697036](https://pubmed.ncbi.nlm.nih.gov/16697036/)]
42. Frezza M, di PC, Pozzato G, Terpin M, Baraona E, Lieber CS. High blood alcohol levels in women. The role of decreased gastric alcohol dehydrogenase activity and first-pass metabolism. *N Engl J Med* 1990 Jan 11;322(2):95-99. [doi: [10.1056/NEJM19901113220205](https://doi.org/10.1056/NEJM19901113220205)] [Medline: [2248624](https://pubmed.ncbi.nlm.nih.gov/2248624/)]
43. Snipes DJ, Benotsch EG. High-risk cocktails and high-risk sex: examining the relation between alcohol mixed with energy drink consumption, sexual behavior, and drug use in college students. *Addict Behav* 2013 Jan;38(1):1418-1423. [doi: [10.1016/j.addbeh.2012.07.011](https://doi.org/10.1016/j.addbeh.2012.07.011)] [Medline: [23006245](https://pubmed.ncbi.nlm.nih.gov/23006245/)]
44. Messman-Moore TL, Ward RM, Zerubavel N. The role of substance use and emotion dysregulation in predicting risk for incapacitated sexual revictimization in women: results of a prospective investigation. *Psychol Addict Behav* 2013 Mar;27(1):125-132. [doi: [10.1037/a0031073](https://doi.org/10.1037/a0031073)] [Medline: [23276308](https://pubmed.ncbi.nlm.nih.gov/23276308/)]

45. Hutton HE, McCaul ME, Chander G, Jenckes MW, Nollen C, Sharp VL, et al. Alcohol use, anal sex, and other risky sexual behaviors among HIV-infected women and men. *AIDS Behav* 2013 Jun;17(5):1694-1704. [doi: [10.1007/s10461-012-0191-4](https://doi.org/10.1007/s10461-012-0191-4)] [Medline: [22566077](https://pubmed.ncbi.nlm.nih.gov/22566077/)]
46. Mah TL, Halperin DT. Concurrent sexual partnerships and the HIV epidemics in Africa: evidence to move forward. *AIDS Behav* 2010 Feb;14(1):11-6; discussion 34. [doi: [10.1007/s10461-008-9433-x](https://doi.org/10.1007/s10461-008-9433-x)] [Medline: [18648926](https://pubmed.ncbi.nlm.nih.gov/18648926/)]
47. Onoya D, Zuma K, Zungu N, Shisana O, Mehlomakhulu V. Determinants of multiple sexual partnerships in South Africa. *J Public Health (Oxf)* 2015 Mar;37(1):97-106. [doi: [10.1093/pubmed/fdu010](https://doi.org/10.1093/pubmed/fdu010)] [Medline: [24639477](https://pubmed.ncbi.nlm.nih.gov/24639477/)]
48. Eaton L, Flisher AJ, Aarø LE. Unsafe sexual behaviour in South African youth. *Soc Sci Med* 2003 Jan;56(1):149-165. [Medline: [12435558](https://pubmed.ncbi.nlm.nih.gov/12435558/)]
49. Pettifor AE, Rees HV, Kleinschmidt I, Steffenson AE, MacPhail C, Hlongwa-Madikizela L, et al. Young people's sexual health in South Africa: HIV prevalence and sexual behaviors from a nationally representative household survey. *AIDS* 2005 Sep 23;19(14):1525-1534. [Medline: [16135907](https://pubmed.ncbi.nlm.nih.gov/16135907/)]
50. 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](https://doi.org/10.1371/journal.pone.0051998)] [Medline: [23284847](https://pubmed.ncbi.nlm.nih.gov/23284847/)]

## Abbreviations

**HIV:** human immunodeficiency virus

**AOR:** adjusted odds ratio

*Edited by G Eysenbach; submitted 09.07.15; peer-reviewed by M Bruening; comments to author 09.10.15; accepted 21.03.16; published 28.09.16.*

*Please cite as:*

*Dietrich JJ, Laher F, Hornschuh S, Nkala B, Chimoyi L, Otwombe K, Kaida A, Gray GE, Miller C*

*Investigating Sociodemographic Factors and HIV Risk Behaviors Associated With Social Networking Among Adolescents in Soweto, South Africa: A Cross-Sectional Survey*

*JMIR Public Health Surveill* 2016;2(2):e154

URL: <http://publichealth.jmir.org/2016/2/e154/>

doi: [10.2196/publichealth.4885](https://doi.org/10.2196/publichealth.4885)

PMID: [27683173](https://pubmed.ncbi.nlm.nih.gov/27683173/)

©Janan Janine Dietrich, Fatima Laher, Stefanie Hornschuh, Busisiwe Nkala, Lucy Chimoyi, Kennedy Otwombe, Angela Kaida, Glenda Elisabeth Gray, Cari Miller. Originally published in *JMIR Public Health and Surveillance* (<http://publichealth.jmir.org>), 28.09.2016. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in *JMIR Public Health and Surveillance*, is properly cited. The complete bibliographic information, a link to the original publication on <http://publichealth.jmir.org>, as well as this copyright and license information must be included.

Original Paper

# A Test of Concept Study of At-Home, Self-Administered HIV Testing With Web-Based Peer Counseling Via Video Chat for Men Who Have Sex With Men

Jessica L Maksut<sup>1</sup>, MA; Lisa A Eaton<sup>1</sup>, PhD; Elizabeth J Siembida<sup>1</sup>, PhD; Daniel D Driffin<sup>2</sup>, BS; Robert Baldwin<sup>2</sup>, BA

<sup>1</sup>Department of Human Development and Family Studies, University of Connecticut, Storrs Mansfield, CT, United States

<sup>2</sup>Institute for Collaboration on Health, Intervention, and Policy, University of Connecticut, Storrs Mansfield, CT, United States

**Corresponding Author:**

Jessica L Maksut, MA

Department of Human Development and Family Studies

University of Connecticut

348 Mansfield Road, Unit 1058

Storrs Mansfield, CT, 06269-1058

United States

Phone: 1 207 590 8207

Fax: 1 860 486 3452

Email: [jessica.maksut@uconn.edu](mailto:jessica.maksut@uconn.edu)

## Abstract

**Background:** Men who have sex with men (MSM), particularly MSM who identify as African-American or Black (BMSM), are the sociodemographic group that is most heavily burdened by the human immunodeficiency virus (HIV) epidemic in the United States. To meet national HIV testing goals, there must be a greater emphasis on novel ways to promote and deliver HIV testing to MSM. Obstacles to standard, clinic-based HIV testing include concerns about stigmatization or recognition at in-person testing sites, as well as the inability to access a testing site due to logistical barriers.

**Objective:** This study examined the feasibility of self-administered, at-home HIV testing with Web-based peer counseling to MSM by using an interactive video chatting method. The aims of this study were to (1) determine whether individuals would participate in at-home HIV testing with video chat-based test counseling with a peer counselor, (2) address logistical barriers to HIV testing that individuals who report risk for HIV transmission may experience, and (3) reduce anticipated HIV stigma, a primary psychosocial barrier to HIV testing.

**Methods:** In response to the gap in HIV testing, a pilot study was developed and implemented via mailed, at-home HIV test kits, accompanied by HIV counseling with a peer counselor via video chat. A total of 20 MSM were enrolled in this test of concept study, 80% of whom identified as BMSM.

**Results:** All participants reported that at-home HIV testing with a peer counseling via video chat was a satisfying experience. The majority of participants (13/18, 72%) said they would prefer for their next HIV testing and counseling experience to be at home with Web-based video chat peer counseling, as opposed to testing in an office or clinic setting. Participants were less likely to report logistical and emotional barriers to HIV testing at the 6-week and 3-month follow-ups.

**Conclusions:** The results of this study suggest that self-administered HIV testing with Web-based peer counseling is feasible and that MSM find it to be a satisfactory means by which they can access their test results. This study can serve as a general guideline for future, larger-scale studies of Web-based HIV test counseling for MSM.

(*JMIR Public Health Surveill* 2016;2(2):e170) doi:[10.2196/publichealth.6377](https://doi.org/10.2196/publichealth.6377)

**KEYWORDS**

HIV prevention; barriers to HIV testing; HIV stigma; MSM; Web-based; Internet; video chat

## Introduction

### Background

Men who have sex with men (MSM), and in particular, MSM who identify as African-American or Black (BMSM), historically have been, and continue to be, the group most heavily impacted by the human immunodeficiency virus (HIV) epidemic in the United States [1,2]. While MSM comprise approximately 2% of the US population, 67% of all new HIV diagnoses in 2014 occurred among men who identify as gay, bisexual, or same gender loving [3]. President Obama's Updated National HIV/AIDS Strategy calls for expanded efforts to prevent HIV by using a combination of effective, evidence-based approaches to improve uptake of HIV prevention and treatment strategies [4]. While there have been advances in improving the adoption of these strategies among MSM [5-8], there continues to be an urgent need for further work that addresses the issue of access to comprehensive HIV care services, including HIV testing, for this population.

For MSM, the most commonly cited barriers to clinic- or office-based HIV testing include logistical concerns (eg, inconvenience of a testing site's hours of operation [9]) and psychosocial issues (eg, fear of discrimination, disclosure of one's sexual identity [10]). Pertinent social determinants of MSM's health, including perceived sexual orientation-based stigmatization by medical professionals, and associated mistrust of the medical community, act as barriers to accessing HIV care, including routine HIV testing and counseling. Indeed, there is evidence that fear of stigmatization may negatively impact patients' health-related quality of life, adherence to treatment regimens, and HIV risk behaviors [11]. In light of the facts that (1) HIV testing is a critical entry point to engagement in HIV care, and (2) multiple barriers discourage MSM from routinely HIV testing, it is worth exploring novel, nonstandard approaches to offering HIV testing services to this population.

In an effort to expand upon current testing approaches, alternative venue HIV prevention outreach efforts that target MSM have emerged. These approaches include behavioral education programs focused on HIV testing uptake, mobile HIV testing in public and semipublic places (eg, parks, streets, community spaces), and linking HIV testing with lesbian, gay, bisexual, and transgender-identified services (eg, community centers) [12]. While beneficial, it is of concern that these approaches may not reach MSM who are not open about their sexual orientation and/or behavior, those who are younger, and those who report less education [13], as well as MSM who would prefer to test in the privacy of their own home [14]. Alternative routes to HIV testing and counseling that are easy to access, offer relative anonymity, can be conducted at home, and offer support by a peer educator may be particularly appealing for MSM.

There exists evidence for the benefits of incorporating Internet delivery as a modality for increasing HIV testing uptake. A recent systematic review of Web-, mobile phone-, and social media-based interventions to address the HIV continuum of care by Muessig et al [15] found that 5 Web-based HIV testing interventions had been published, and that an additional 8 were

currently in development. Included in the review was a social media-based HIV testing intervention, Project Hope, which randomly assigned peer leaders to deliver either HIV information and promotion of free home-based HIV tests (ie, intervention) or general health information (ie, control) via social media [16]. Young et al [16] found that the intervention group participants were more likely to request home-based HIV testing kits, and take and return test kits than control participants. A number of additional Web-based studies [17-20] with similar formats to those mentioned above (Web-, mobile phone-, and social media-based) aimed at promoting HIV testing exist, but according to Schnall et al [21], they vary greatly with respect to study quality, and the majority, unlike Project Hope, aim to increase rates of clinic-based HIV testing. One such example is a Web-based HIV testing intervention by Bauermeister and colleagues [22], which offered tailored, motivational content around HIV/sexually transmitted infections (STIs) testing as well as a testing locator to MSM. While Bauermeister et al [22] found that those who received the intervention (compared with those who received the testing locator only) were more likely to visit an HIV/STI testing clinic, it is possible that some participants still may have not tested due to barriers to clinic-based testing. Like Bauermeister et al, Zou et al [23] found that active (eg, instant messaging, chat rooms, mobile phones, email) promotion methods were more effective at promoting HIV testing than passive (eg, posters) methods. Taken together, previous Web-based studies of HIV testing point to the usefulness of the Internet's reach, as well as the importance of active methods to promote HIV testing; however, most studies were structured around motivating participants to access clinic-based testing rather than bringing testing directly to individuals by mailing test kits.

Now that there are at-home HIV test kits available, more should be done to understand how the Internet can augment and improve the experience of HIV testing at home. According to Katz et al [24], who conducted a study on the acceptability and ease of use of at-home HIV test kits among MSM, 96% of MSM found the test kits very easy to use, while the remaining participants found them somewhat easy to use. The majority of participants in Katz et al's study [24] requested additional kits after the first. Choko et al [25] found in their study on the uptake and accuracy of at-home, oral HIV test kits that accuracy was good, though approximately 10% of participants made minor procedural errors and 10% reported wanting or needing supervisory support. Based on the previously mentioned review of prior Web-based HIV testing interventions, as well as the evidence of the accuracy, acceptability, and ease of use of at-home testing, offering Web-based HIV counseling to accompany at-home HIV testing experiences appears to be a relatively untapped area for intervention development.

### Study Objectives

The current study examined the feasibility of self-administered, at-home HIV testing with Web-based peer counseling to MSM by using an interactive video chatting method (eg, Skype, Google Hangouts, FaceTime). The primary aim was to conduct a test of concept study to determine whether individuals would participate in at-home HIV testing with video chat-based test counseling with a peer counselor. A secondary aim was to gather



preliminary evidence that this testing modality could (1) address logistical barriers to HIV testing that individuals who report risk for HIV transmission may experience, and (2) reduce anticipated HIV stigma, a primary psychosocial barrier to HIV testing. Pre- and posttesting assessments were used to evaluate changes in variables of interest, including barriers to HIV testing and anticipated HIV stigma.

## Methods

### Participants

Recruitment occurred through word of mouth, phone call-ins, and Web-based advertisements on dating websites for gay and bisexual men. Recruitment targeted BMSM in particular, though MSM who were not Black/African-American but had heard about the study via word of mouth and expressed interest were still invited to participate. Participants were recruited over a 3-month period from January to March 2015. To be eligible, participants reported having condomless anal sex with a man in the past 6 months; having an HIV negative or unknown status; being at least 18 years of age; having access to a computer, tablet, mobile phone, or other device with video chatting capabilities and Internet/mobile service; and agreed to receive an HIV test kit via mail. Twenty men in the Atlanta, GA metropolitan area provided informed consent and were enrolled. Of those, 18 participants completed the 6-week and 3-month follow-up assessments. Participants were provided monetary incentives for completing survey assessments, specifically US \$35 for completing the baseline appointment and US \$25 for each follow-up appointment. All study procedures were approved by the University of Connecticut institutional review board.

### Sequence of Study Events

First, participants who expressed interest in the study completed a consent appointment via a phone call with the peer counselor. The peer counselor shared many of the same demographic characteristics with the majority of the study participants (eg, race/ethnicity, approximate age, gender, sexual orientation, and geographic location) and had the state of Georgia pre- and posttest HIV counseling certification. During the consent appointment, participants were explained the study procedures, read the consent form, were invited to ask questions, and provided consent via a mobile consenting process where participants typed their name into an electronic consent form that was emailed to them. Upon completing the consent procedures, participants were scheduled for their baseline appointment and were mailed an HIV test to a location of their choice. The arrival of the HIV test kit and the baseline appointment were both scheduled for the same day. The type of HIV test that was mailed to participants was the ORAQuick ADVANCE Rapid HIV-1/2 Antibody Test by OraSure Technologies, Inc. This test was selected due to its Clinical Laboratory Improvement Amendments waiver, meaning that it is approved for use outside of laboratory settings.

During the baseline appointment, which occurred via video chatting (eg, Skype, FaceTime), participants engaged with the peer counselor in pretest HIV counseling, self-administration of the HIV test (with guidance from the peer counselor), a

survey assessment using audio computer-assisted self-interview software (which was completed during the running of the HIV test and took approximately 25 minutes to complete), and posttest HIV counseling. All study procedures, including the self-administration of the HIV test and the baseline survey assessment, were completed during the video chat sessions. The first study appointment took approximately 45 minutes to complete.

The peer-delivered counseling was comprised of content with multiple themes. With the peer counselor's guidance and taking a harm reduction approach, participants worked through a practical and tailored sexual reduction plan based on their reported HIV risk-taking behaviors. Moreover, substance use (ie, alcohol and drug use) in the context of sexual risk taking was assessed and incorporated into the counseling session. Social, emotional, and structural barriers, including anticipated HIV stigma, to engaging in routine HIV testing were evaluated, discussed, and problem solved. In addition, proper referrals to additional services, as needed, were provided by the peer counselor.

The follow-up appointments were completed via a phone call with noncounseling staff members. During the follow-up appointments, the assessment questions and response sets were read to participants and participants provided their answers. Each follow-up appointment took approximately 25 minutes to complete.

### Measures

#### *Sociodemographic Characteristics*

Participants were asked to report on their age, years of education, employment status, marital status, income level, race/ethnicity, and sexual orientation (ie, whether they identified as same gender loving/gay, bisexual, or heterosexual). Further, participants were asked to report whether they had ever taken an HIV test, and if so, the date of their last HIV test (reported in number of months between date of appointment and date of last test).

#### *Feasibility, Acceptability, and HIV Testing Outcomes*

Feasibility of recruiting for the study was assessed by tracking the number of men approached, the number who agreed to participate, and the number deemed eligible. Participants' rates of retention were based on whether they completed the 6-week and the 3-month follow-up sessions. At the end of the study, participants were asked to answer questions regarding their satisfaction with their experience conducting HIV testing via video chatting. Items included: "Would you like to test for HIV at home with video chat-based peer counseling again?" and "Would you recommend testing for HIV at home with video chat-based peer counseling to one of your friends?" Items were dichotomous, 0=no, 1=yes. Also included was "Would you prefer your next HIV test to be in person at an office or at home with peer counseling via video chat?" Response options included "an office" or "at home with video chat-based peer counseling." Further, participants were asked to rate how satisfied they were with their experience testing at home with video chat-based peer counseling. This item ranged from 0 (not satisfied) to 2 (very satisfied). Additionally, HIV test results, as well as the



location of the participants during testing (eg, dorm room, bedroom) were reported.

### **Barriers to HIV Testing**

Participants' barriers to HIV testing—including structural barriers, such as lack of transportation and distance to testing site, as well as concerns regarding confidentiality and fear of testing HIV positive—were assessed using the Barriers to HIV Testing scale [26]. Items included: “The testing site is too far away” and “I am concerned about how I will be treated at the testing site.” Items were rated on a Likert-type scale from 1 (strongly disagree) to 6 (strongly agree).

### **Anticipated HIV Stigma**

We assessed the extent to which participants anticipated negative intra- and interpersonal consequences of testing HIV positive in the future using 5 items adapted from the Anticipated HIV Stigma scale [27]. Items included: “I would feel I let myself down if I ever got infected with HIV,” “If I got infected with HIV, no one would date me,” and “I would feel I were not as good a person as others if I got HIV.” All items were rated on a Likert-type scale from 1 (strongly disagree) to 6 (strongly agree).

### **Data Analysis**

Due to the small sample size, both *t* tests and effect sizes—specifically Cohen *d* [28]—were used to assess treatment effects at the 6-week and 3-month posttest assessments. Cohen *d* was calculated using pre- and posttest scores (at both the 6-week and 3-month assessments) for individual anticipated stigma and barriers to treatment items. This approach was done primarily to understand which barriers to HIV testing, specifically, this modality impacts, and to determine whether anticipated stigma from various sources (eg, family, friends, romantic or sexual partners) are differentially affected by at-home, self-administered HIV testing with video chat-based peer counseling.

In prior research of this nature, Cohen *d* values larger than 0.30 (or  $-0.30$ ) have been considered to be medium effect sizes and indicate potential change between baseline and follow-ups [29]. Results of significant *t* tests, as well as effect sizes above 0.30 (or below  $-0.30$ ), are reported for the 6-week and 3-month outcomes in the proceeding text.

## **Results**

### **Sociodemographic Characteristics**

Seventy percent (14/20) of men identified as gay, homosexual, or same gender loving, and the remaining 30% (6/20) identified as bisexual. Eighty percent (16/20) identified as Black, non-Hispanic or Latino, 15% (3/20) identified as White, non-Hispanic or Latino, and the remaining individual (1/20, 5%) identified as White, Hispanic or Latino. On average, participants were approximately 28 years of age. Of participants, 100% (20/20) were single, and the majority of participants

reported that they were currently working (16/20, 80%). The average length of time since participants' last HIV test was approximately 12 months ( $SD=22.9$ ). Of participants, 85% (17/20) had taken at least 1 HIV test in the past (Table 1).

### **Feasibility, Acceptability, and HIV Testing Outcomes**

Fifty potential participants were screened and described the study opportunity; 15 individuals were screened out based on eligibility criteria, and 7 declined to participate. Twenty-three participants (23/50, 46%) were interested and agreed to participate. Of the 23 men who agreed, 87% (20/23) completed the HIV testing appointment and 90% (18/20) were retained at follow-ups.

Participants' responses to the satisfaction items demonstrated that all participants found at-home HIV testing with video chat-based peer counseling to be satisfying. All participants reported that they would like to participate in at-home HIV testing with peer counseling via video chat in the future (18/18, 100%) and that they would recommend this modality to one of their friends (18/18, 100%). Further, 72% of participants (13/18) said that they would prefer for their next HIV test to be self-administered at home with counseling from a peer via video chat, as opposed to in a clinic or an office setting (Table 2).

Participants were able to take their HIV tests in a variety of locations, including in their homes (12/20, 60%), in their garages (1/20, 5%), on their porches (1/20, 5%), in dorm rooms (2/20, 10%), at friends' houses (2/20, 10%), at work (1/20, 5%), and in their car (1/20, 5%). All participants tested HIV negative (20/20, 100%) during the counseling session (Table 3). Further, qualitative observations highlighted that participants frequently reported the importance of being able to administer at-home HIV tests and engage in HIV test counseling via video chat with a peer counselor in a variety of locations. Overwhelmingly, participants reported that they did not want others to know they were taking an HIV test, and that this HIV testing and counseling methodology allowed for greater flexibility in testing location and for control regarding their privacy (ie, who would know that they were testing).

### **Barriers to HIV Testing**

Participants were less concerned about how they would be treated by people at the testing site at the 6-week posttest ( $d=0.34$ ) than at baseline. Similarly, a number of medium effect sizes emerged at the 3-month follow-up, including less concern regarding (1) lack of transportation ( $d=0.31$ ), (2) testing sites being too far away ( $d=0.32$ ), (3) how they will be treated at the testing site ( $d=0.37$ ), and (4) finding out the results of their HIV tests ( $d=0.37$ ). For each of these items, participants' barriers were reduced from baseline to the 3-month posttest (Table 4). A second qualitative observation included noting that participants reported the video chat-based peer counseling component of their HIV testing experiences as an important and helpful opportunity to troubleshoot barriers to HIV testing in the future.

**Table 1.** Demographic characteristics of men who have sex with men (MSM) recruited from the Atlanta, GA, area for HIV testing via video chat.

Demographic information	Mean (range) or n	SD or %
Age, mean (range), SD	28.05 (20-44)	6.80
Education, mean (range), SD	2.15 (1-4)	1.23
Income, mean (range), SD	3.20 (1-7)	1.80
<b>Sexual orientation, n, %</b>		
Gay/homosexual/same gender loving	14	70
Bisexual	6	30
<b>Race/ethnicity, n, %</b>		
Black, non-Hispanic or Latino	16	80
White, non-Hispanic or Latino	3	15
White, Hispanic or Latino	1	5
<b>Gender, n, %</b>		
Male	19	95
Transgender female	1	5
<b>Marital status, n, %</b>		
Single	20	100
Domestic partnership/Civil union	0	0
<b>Employment status, n, %</b>		
Unemployed	4	20
Working	16	80
Time since last HIV test in months, mean (range), SD	12.31 (1.07-95.40)	22.93
<b>Have you ever taken an HIV test in the past? n, %</b>		
Yes	17	85
No	3	15.0

**Table 2.** Participants' satisfaction with at-home HIV testing with peer counseling via video chat at the 3-month follow-up.

Patient satisfaction questions	Mean (range) or n	SD or %
How satisfied were you with at-home HIV testing with video chat-based peer counseling? mean (range), SD	1.89 (0-2)	0.32
<b>Would you like to test for HIV at home with video chat-based peer counseling again? mean (range), SD</b>		
Yes	18	100
No	0	0
<b>Would you recommend testing for HIV at home with video chat-based peer counseling to one of your friends? mean (range), SD</b>		
Yes	18	100
No	0	0
<b>Would you prefer your next HIV test to be in person at an office or at home with video chat-based peer counseling? mean (range), SD</b>		
At an office	5	28
At home with video chat-based peer counseling	13	72

**Table 3.** Participants' HIV testing outcomes from at-home HIV testing with HIV counseling via video chat.

Testing outcomes	n (%)
<b>Result of HIV test</b>	
HIV negative	20 (100)
HIV positive	0 (0)
<b>Location of HIV test</b>	
Home, general	12 (60)
Home, in garage	1 (5)
Home, on porch	1 (5)
In dorm room	2 (10)
At a friend's house	2 (10)
Inside of car	1 (5)
At work	1 (5)

**Table 4.** Means and standard deviations of the anticipated stigma and barriers to testing items at baseline, 6-week, and 3-month follow-ups.

Variable	Baseline		6-week post		3-month post		6-week effect size	3-month effect size	P value
	M	SD	M	SD	M	SD			
<b>Barriers to testing items</b>									
I don't have transportation to site.	1.61	1.50	1.56	1.34	1.22	0.94	0.04	0.31	.13
I don't have enough time.	1.94	1.80	1.78	1.35	1.67	1.41	0.10	0.17	.57
The testing site is too far away.	1.72	1.71	1.50	1.04	1.28	0.96	0.16	0.32	.15
I don't know where to go for testing.	1.39	1.24	1.33	0.97	1.72	1.67	0.05	-0.22	.39
I am concerned about how I will be treated by people at the testing site.	2.50	2.07	1.89	1.53	1.83	1.52	0.34	0.37	.19
I have had a bad HIV testing experience in the past.	1.72	1.64	1.39	1.15	1.61	1.50	0.23	0.07	.76
I can't afford treatment, so why get tested?	1.39	1.24	1.33	1.03	1.33	1.03	0.05	0.05	.75
I don't want to know the results.	2.11	2.03	1.72	1.57	1.50	1.20	0.21	0.37	.15
People might recognize me at the testing site.	2.39	2.20	2.89	2.08	2.00	1.72	-0.23	0.20	.42
I am worried about my health information being kept confidential.	2.00	1.97	2.50	1.86	1.72	1.23	-0.26	0.17	.58
<b>Anticipated stigma items</b>									
I would let myself down if I ever got infected with HIV.	5.17	1.30	5.56	0.71	5.22	1.31	-0.37	-0.04	.90
I would let my family and friends down if I ever got infected with HIV.	3.78	1.99	4.50	1.82	4.67	1.88	-0.38	-0.46	.04
If I got infected with HIV, no one would date me.	2.17	1.69	2.33	1.68	2.33	1.94	-0.09	-0.09	.66
If I got infected with HIV, men would not want to have sex with me.	3.11	1.97	2.33	1.50	2.28	1.81	0.44	0.44	.07
I would feel I were not as good a person if I got HIV.	2.28	1.71	2.78	1.93	2.33	1.78	-0.27	-0.03	.90

## Anticipated HIV Stigma

An effect size of  $d=-0.37$  for the “I would let myself down if I ever got infected with HIV” item was noted, indicating that, at the 6-week follow-up, participants were more likely to feel that they would let themselves down if they ever got infected with HIV. Moreover, for the item “I would let my family and friends down if I ever got infected with HIV,”  $d=-0.38$  at the 6-week follow-up and  $d=-0.46$  at the 3-month follow-up, suggesting that participants were more likely to feel that they would let their family and friends down if they got infected with HIV at follow-ups than baseline. Further, there was a significantly different change in mean scores from pre- to 3-month posttest for this item,  $t_{(17)}=.16$ ,  $P=.04$ . For the item “If I got infected with HIV, men would not want to have sex with me,”  $d=0.44$  at both the 6-week and 3-month follow-ups, meaning that participants were less likely to feel that getting infected with HIV would result in men not wanting to have sex with them at the follow-ups than at baseline (Table 4).

## Discussion

### Principal Findings

The results of this test of concept study suggest that at-home, self-administered HIV testing with peer counseling via video chat is well received and has the potential to benefit participants and to serve as a viable and novel HIV testing and counseling approach. A substantial number of the men who were asked to participate in this study did so and those who did participate found this new form of HIV testing and counseling to be a highly satisfying experience. Further, a majority of participants reported preferring this approach to testing over testing in an office setting in the future. All participants reported that they would like to test for HIV using this method again (ie, at home testing with peer counseling via video chat) and that they would recommend to their friends this method of HIV testing and counseling. These findings suggest that this HIV testing and counseling approach is well-received by MSM, which points to its potential to make an impact on improving HIV testing uptake and frequency among this higher risk population [30]. These findings warrant support for further study of at-home HIV testing and peer counseling via video chat.

The benefits of administering Web-based HIV counseling are numerous, including being less time-intensive and more convenient, as well as the relative anonymity for participants and the flexibility with which they can choose the location to take their tests [7,31]. Some participants in this study tested in places other than their homes, including in their cars, at their workplaces, in their garages, and at their friends' homes. However, it is important to note that all participants were encouraged to find a safe and comfortable place to test where their privacy could be protected. In cases where participants lived with relatives or friends, the ability to take their tests outside of the home was beneficial for maintaining their privacy and the Web-based, video chatting format of the counseling did well to allow for this important flexibility.

Interestingly, a number of moderate effect sizes were found, including a notable decrease in barriers to testing, namely

logistical (eg, transportation) and psychosocial (eg, fear of discrimination or being recognized at the testing site) barriers [32-34]. The present study is a pilot study with a small sample size, and, as such, these results should be interpreted with caution; however, they provide preliminary support for findings that could be further evaluated by larger scale studies. In previous research, these barriers have been identified as places of weakness in the HIV continuum of care, particularly for the most vulnerable communities (eg, BMSM), for whom the burden of undetected HIV infection strongly persists [2,35,36]. Arguably, these barriers may have been lessened because of the Web-based nature of this study. It is possible that by introducing participants to an alternative route to HIV testing (ie, testing at home with video chat-based peer counseling) previously held logistical or stigma-related concerns about accessing HIV testing were abated. Further, the peer counselors worked to problem solve barriers to seeking out local HIV testing sites during the counseling sessions.

This study found that anticipated HIV stigma both increased and decreased from pre- to posttest. While participants were more likely to feel that they would let their family and friends down if they were to be infected with HIV in the future, they were less likely to report that men would not want to have sex with them if they became infected with HIV. It is possible that exposure to new HIV knowledge or having conversations with a peer educator about HIV stimulated newfound anxiety among the participants regarding their HIV vulnerability and/or having to manage telling family and friends about an HIV diagnosis. As part of the sexual risk reduction portion of the counseling experience, however, participants conversed with the peer educator about partner selection strategies, which may have led to the decreased scores related to HIV stigma from romantic or sexual partners at the 6-week and 3-month posttests. This decrease in anticipated HIV stigma is important because research has indicated that HIV stigma negatively impacts the health of MSM by limiting access to health care, discouraging routine HIV testing, and contributing to stress, social isolation, and risky sex behaviors [37-40]. Future testing modalities of this nature should include discussions concerning not only sexual and romantic partners, but family and friends as well [41].

### Strengths and Limitations

To our knowledge, this is the first test of concept of at-home HIV testing with video chat-based peer counseling for MSM. This study adds to the extant literature by identifying the feasibility of HIV test counseling via video chat. One strength of the current study's sample is that the majority of participants identified as African-American or Black. The need for novel HIV testing strategies for BMSM is urgent given the alarmingly high rates of HIV among this community, and the presently assessed modality—at-home, self-administered HIV testing with peer counseling via video chat—was well-received by BMSM. However, it is also worth noting that the study used surveys, which relied on self-reporting of potentially sensitive experiences and behaviors and, therefore, may be prone to bias. Additionally, social desirability bias may have impacted the results of the study by affecting study satisfaction-related outcomes, among other study data points. Further, participants were asked whether they wanted their next HIV test to be at an

office or at home with video chat-based peer counseling, but did not provide participants with other testing venue options to choose from (eg, community clinic). Another important limitation to the study is the cost associated with at-home HIV test kits. Given the expense of personally purchased, at-home HIV test kits, community-based organizations must provide test kits for uptake of the proposed testing strategy.

Protocols were developed and put in place to support participants testing HIV positive during the HIV testing piece of the video chat-based peer counseling. Specifically, our procedures for providing participants who test HIV positive appropriate support and linkage to care included the following multiple steps: (1) the provision of Center for Disease Control-based post-HIV test counseling for individuals testing HIV positive; (2) evaluation of mental health status and referral to immediate or delayed care as determined by the counselor; (3) linkage to long-term HIV care organization of the counselor's referral and/or the participant's choice; (4) the arrangement of initial HIV care appointment by the counselor; (5) review with participant, as needed, services provided by selected HIV care organization (eg, health insurance assistance, substance use and mental health treatment, housing, transportation, support groups, medication counseling, transportation, etc); (6) re-contacting the participant 2 days, 2 weeks, and 2 months posttesting appointment to ensure proper follow-up care and to problem

solve obstacles; and (7) report test results to the State Health Department (additional details on at-home HIV testing protocol can be obtained by contacting the primary author of this manuscript). Though we had these procedures in place, all participants in our sample tested HIV negative. It is possible that the study results were impacted by this particular outcome. Future studies are needed to assess the successes and challenges of linking individuals who test HIV positive to care.

## Conclusions

By assessing the acceptance and feasibility of at-home, self-administered HIV testing with video chat-based peer counseling among this population, this study gains a key insight as to how such an intervention can be best delivered to individuals in need of HIV testing and counseling, but who may not have access to or who may not prefer more common routes to testing and other services, including clinic-based services. The study demonstrated feasibility, and the participants were satisfied with their experiences, indicating not only that this testing and counseling modality could likely be replicated, but also that a Web-based, video chat approach to HIV counseling is acceptable to participants. Given the paucity of alternative venue HIV testing locations available to MSM, as well as the number of MSM who are not able or willing to access services, this study is an important step toward filling an unmet need.

## Acknowledgments

This project was supported by National Institute of Mental Health Grant R01MH094230.

## Conflicts of Interest

None declared.

## References

1. Rosenberg ES, Grey JA, Sanchez TH, Sullivan PS. Rates of prevalent HIV infection, prevalent diagnoses, and new diagnoses among men who have sex with men in US states, metropolitan statistical areas, and counties. *JMIR Public Health Surveill* 2016;2:e22 [FREE Full text] [doi: [10.2196/publichealth.5684](https://doi.org/10.2196/publichealth.5684)] [Medline: [27244769](https://pubmed.ncbi.nlm.nih.gov/27244769/)]
2. Mannheimer SB, Wang L, Wilton L, Van TH, Del RC, Buchbinder S, et al. Infrequent HIV testing and late HIV diagnosis are common among a cohort of black men who have sex with men in 6 US cities. *J Acquir Immune Defic Syndr* 2014;67:438-445 [FREE Full text] [doi: [10.1097/QAI.0000000000000334](https://doi.org/10.1097/QAI.0000000000000334)] [Medline: [25197830](https://pubmed.ncbi.nlm.nih.gov/25197830/)]
3. Purcell D, Johnson C, 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. [doi: [10.2174/1874613601206010098](https://doi.org/10.2174/1874613601206010098)] [Medline: [23049658](https://pubmed.ncbi.nlm.nih.gov/23049658/)]
4. Office of National AIDS Policy. National HIV/AIDS Strategy: Updated to 2020. 2015. URL: [https://www.whitehouse.gov/sites/default/files/docs/national\\_hiv\\_aids\\_strategy\\_update\\_2020.pdf](https://www.whitehouse.gov/sites/default/files/docs/national_hiv_aids_strategy_update_2020.pdf) [accessed 2016-11-28] [WebCite Cache ID 6mMYOICqF]
5. Anaya HD, Hoang T, Golden JF, Goetz MB, Gifford A, Bowman C, et al. Improving HIV screening and receipt of results by nurse-initiated streamlined counseling and rapid testing. *J Gen Intern Med* 2008;23:800-807 [FREE Full text] [doi: [10.1007/s11606-008-0617-x](https://doi.org/10.1007/s11606-008-0617-x)] [Medline: [18421508](https://pubmed.ncbi.nlm.nih.gov/18421508/)]
6. Metcalf C, Malotte C, Douglas JJ, Paul S, Dillon B, Cross H, et al. Efficacy of a booster counseling session 6 months after HIV testing and counseling: A randomized, controlled trial (RESPECT-2). *Sex Transm Dis* 2005;32:123-129. [Medline: [15668620](https://pubmed.ncbi.nlm.nih.gov/15668620/)]
7. Noar S, Black H, Pierce L. Efficacy of computer technology-based HIV prevention interventions: a meta-analysis. *AIDS* 2009;23:107-115. [doi: [10.1097/QAD.0b013e32831c5500](https://doi.org/10.1097/QAD.0b013e32831c5500)] [Medline: [19050392](https://pubmed.ncbi.nlm.nih.gov/19050392/)]
8. Rhodes SD, Vissman AT, Stowers J, Miller C, McCoy TP, Hergenrather KC, et al. A CBPR partnership increases HIV testing among men who have sex with men (MSM): outcome findings from a pilot test of the CyBER/testing internet intervention. *Health Educ Behav* 2011;38:311-320 [FREE Full text] [doi: [10.1177/1090198110379572](https://doi.org/10.1177/1090198110379572)] [Medline: [21393625](https://pubmed.ncbi.nlm.nih.gov/21393625/)]



9. 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;41:218-224. [Medline: [16394855](#)]
10. Spielberg F, Branson BM, Goldbaum GM, Lockhart D, Kurth A, Celum CL, et al. Overcoming barriers to HIV testing: preferences for new strategies among clients of a needle exchange, a sexually transmitted disease clinic, and sex venues for men who have sex with men. *J Acquir Immune Defic Syndr* 2003;32:318-327. [Medline: [12626893](#)]
11. Whetten K, Reif S, Whetten R, Murphy-McMillan LK. Trauma, mental health, distrust, and stigma among HIV-positive persons: implications for effective care. *Psychosom Med* 2008;70:531-538. [doi: [10.1097/PSY.0b013e31817749dc](#)] [Medline: [18541904](#)]
12. Warwick I, Douglas N, Aggleton P, Boyce P. Context matters: the educational potential of gay bars revisited. *AIDS Educ Prev* 2003;15:320-333. [Medline: [14516017](#)]
13. Wendell D, Cohen D, LeSage D, Farley T. Street outreach for HIV prevention: effectiveness of a state-wide programme. *Int J STD AIDS* 2003;14:334-340. [doi: [10.1258/095646203321605549](#)] [Medline: [12803941](#)]
14. Orellana ER, Picciano JF, Roffman RA, Swanson F, Kalichman SC. Correlates of nonparticipation in an HIV prevention program for MSM. *AIDS Educ Prev* 2006;18:348-361. [doi: [10.1521/aeap.2006.18.4.348](#)] [Medline: [16961451](#)]
15. Muessig KE, Nekkanti M, Bauermeister J, Bull S, Hightow-Weidman LB. A systematic review of recent smartphone, Internet and Web 2.0 interventions to address the HIV continuum of care. *Curr HIV/AIDS Rep* 2015;12:173-190 [FREE Full text] [doi: [10.1007/s11904-014-0239-3](#)] [Medline: [25626718](#)]
16. Young S, Holloway I, Jaganath D, Rice E, Westmoreland D, Project C. Online social network changes in an HIV prevention randomized controlled trial for African American and Latino men who have sex with men. *Am J Public Health* 2014;104:1707-1712. [doi: [10.2105/AJPH.2014.301992](#)] [Medline: [25033137](#)]
17. Blas MM, Alva IE, Carcamo CP, Cabello R, Goodreau SM, Kimball AM, et al. Effect of an online video-based intervention to increase HIV testing in men who have sex with men in Peru. *PLoS One* 2010;5:e10448 [FREE Full text] [doi: [10.1371/journal.pone.0010448](#)] [Medline: [20454667](#)]
18. Bourne C, Knight V, Guy R, Wand H, Lu H, McNulty A. Short message service reminder intervention doubles sexually transmitted infection/HIV re-testing rates among men who have sex with men. *Sex Transm Infect* 2011;87:229-231. [doi: [10.1136/sti.2010.048397](#)] [Medline: [21296796](#)]
19. Ko N, Hsieh C, Wang M, Lee C, Chen C, Chung A, et al. Effects of Internet popular opinion leaders (iPOL) among Internet-using men who have sex with men. *J Med Internet Res* 2013;15:e40 [FREE Full text] [doi: [10.2196/jmir.2264](#)] [Medline: [23439583](#)]
20. Rhodes SD, Vissman AT, Stowers J, Miller C, McCoy TP, Hergenrather KC, et al. A CBPR partnership increases HIV testing among men who have sex with men (MSM): outcome findings from a pilot test of the CyBER/testing internet intervention. *Health Educ Behav* 2011;38:311-320 [FREE Full text] [doi: [10.1177/1090198110379572](#)] [Medline: [21393625](#)]
21. 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:e134. [doi: [10.2196/jmir.3393](#)] [Medline: [24862459](#)]
22. Bauermeister J. Get connected: a tailored web intervention for young MSM. 2014 Presented at: TH Live Conference; April 6-8, 2014; San Francisco, CA URL: <http://www.slideshare.net/isisinc/get-connected-a-tailored-web-intervention-for-young-msm>
23. Zou H, Wu Z, Yu J, Li M, Ablimit M, Li F, et al. Internet-facilitated, voluntary counseling and testing (VCT) clinic-based HIV testing among men who have sex with men in China. *PLoS One* 2013;8:e51919 [FREE Full text] [doi: [10.1371/journal.pone.0051919](#)] [Medline: [23418417](#)]
24. Katz D, Golden M, Hughes J, Farquhar C, Stekler J. Acceptability and ease of use of home self-testing for HIV among men who have sex with men. *Int J STD AIDS* 2016;27:531-536. [doi: [10.1177/0956462415586676](#)]
25. Choko A, Desmond N, Webb E, Chavula K, Napierala-Mavedzenge S, Gaydos C, et al. The uptake and accuracy of oral kits for HIV self-testing in high HIV prevalence setting: a cross-sectional feasibility study in Blantyre, Malawi. *PLoS Med* 2011;8:e1001102. [doi: [10.1371/journal.pmed.1001102](#)] [Medline: [21990966](#)]
26. Awad GH, Sagrestano LM, Kittleson MJ, Sarvela PD. Development of a measure of barriers to HIV testing among individuals at high risk. *AIDS Educ Prev* 2004;16:115-125. [Medline: [15134120](#)]
27. Starks TJ, Rendina HJ, Breslow AS, Parsons JT, Golub SA. The psychological cost of anticipating HIV stigma for HIV-negative gay and bisexual men. *AIDS Behav* 2013;17:2732-2741. [doi: [10.1007/s10461-013-0425-0](#)] [Medline: [23420102](#)]
28. Cohen J. A power primer. *Psychol Bull* 1992;112:155-159. [Medline: [19565683](#)]
29. Payne K, Marcus D. The efficacy of group psychotherapy for older adult clients: a meta-analysis. *Group Dyn* 2008;12:268-278.
30. Conway DP, Guy R, Davies SC, Couldwell DL, McNulty A, Smith DE, Sydney Rapid HIV Test Study. Rapid HIV testing is highly acceptable and preferred among high-risk gay and bisexual men after implementation in Sydney sexual health clinics. *PLoS One* 2015;10:e0123814 [FREE Full text] [doi: [10.1371/journal.pone.0123814](#)] [Medline: [25898140](#)]
31. Angotti N, Bula A, Gaydos L, Kimchi EZ, Thornton RL, Yeatman SE. Increasing the acceptability of HIV counseling and testing with three C's: convenience, confidentiality and credibility. *Soc Sci Med* 2009;68:2263-2270 [FREE Full text] [doi: [10.1016/j.socscimed.2009.02.041](#)] [Medline: [19375208](#)]

32. Obermeyer C, Osborn M. The utilization of testing and counseling for HIV: a review of the social and behavioral evidence. *Am J Public Health* 2007;97:1762-1774. [doi: [10.2105/AJPH.2006.096263](https://doi.org/10.2105/AJPH.2006.096263)] [Medline: [17761565](https://pubmed.ncbi.nlm.nih.gov/17761565/)]
33. Schwarcz S, Richards TA, Frank H, Wenzel C, Hsu LC, Chin CJ, et al. Identifying barriers to HIV testing: personal and contextual factors associated with late HIV testing. *AIDS Care* 2011;23:892-900. [doi: [10.1080/09540121.2010.534436](https://doi.org/10.1080/09540121.2010.534436)] [Medline: [21424942](https://pubmed.ncbi.nlm.nih.gov/21424942/)]
34. Sutton M, Anthony M, Vila C, McLellan-Lemal E, Weidle PJ. HIV testing and HIV/AIDS treatment services in rural counties in 10 southern states: service provider perspectives. *J Rural Health* 2010;26:240-247. [doi: [10.1111/j.1748-0361.2010.00284.x](https://doi.org/10.1111/j.1748-0361.2010.00284.x)] [Medline: [20633092](https://pubmed.ncbi.nlm.nih.gov/20633092/)]
35. Paltiel AD, Weinstein MC, Kimmel AD, Seage GR, Losina E, Zhang H, et al. Expanded screening for HIV in the United States--an analysis of cost-effectiveness. *N Engl J Med* 2005;352:586-595. [doi: [10.1056/NEJMsa042088](https://doi.org/10.1056/NEJMsa042088)] [Medline: [15703423](https://pubmed.ncbi.nlm.nih.gov/15703423/)]
36. Washington T, D'Anna L, Meyer-Adams N, Malotte C. From their voices: barriers to HIV testing among black men who have sex with men remain. *Healthcare* 2015;3:933-947. [doi: [10.3390/healthcare3040933](https://doi.org/10.3390/healthcare3040933)]
37. CDC. CDC Fact Sheet: HIV among gay and bisexual men. 2015. URL: <https://www.cdc.gov/nchhstp/newsroom/docs/factsheets/cdc-msm-508.pdf> [accessed 2016-11-28] [WebCite Cache ID 6mMav0B0A]
38. Chesney M, Smith A. Critical delays in HIV testing and care the potential role of stigma. *Am Behav Sci* 1999;42:1162-1174. [doi: [10.1177/00027649921954822](https://doi.org/10.1177/00027649921954822)]
39. Li X, Lu H, Ma X, Sun Y, He X, Li C, et al. HIV/AIDS-related stigmatizing and discriminatory attitudes and recent HIV testing among men who have sex with men in Beijing. *AIDS Behav* 2012;16:499-507 [FREE Full text] [doi: [10.1007/s10461-012-0161-x](https://doi.org/10.1007/s10461-012-0161-x)] [Medline: [22350831](https://pubmed.ncbi.nlm.nih.gov/22350831/)]
40. Parker R, Aggleton P. HIV and AIDS-related stigma and discrimination: a conceptual framework and implications for action. *Soc Sci Med* 2003;57:13-24. [Medline: [12753813](https://pubmed.ncbi.nlm.nih.gov/12753813/)]
41. Bogart L, Cowgill B, Kennedy D, Ryan G, Murphy D, Elijah J, et al. HIV-related stigma among people with HIV and their families: a qualitative analysis. *AIDS Behav* 2008;12:244-254. [doi: [10.1007/s10461-007-9231-x](https://doi.org/10.1007/s10461-007-9231-x)] [Medline: [17458691](https://pubmed.ncbi.nlm.nih.gov/17458691/)]

## Abbreviations

**BMSM:** men who have sex with men who identify as African-American or Black

**HIV:** human immunodeficiency virus

**MSM:** men who have sex with men

**STI(s):** sexually transmitted infection(s)

*Edited by T Sanchez; submitted 20.07.16; peer-reviewed by MA Chaisson, TS Hottes; comments to author 26.08.16; revised version received 18.10.16; accepted 07.11.16; published 14.12.16.*

*Please cite as:*

*Maksut JL, Eaton LA, Siembida EJ, Driffin DD, Baldwin R*

*A Test of Concept Study of At-Home, Self-Administered HIV Testing With Web-Based Peer Counseling Via Video Chat for Men Who Have Sex With Men*

*JMIR Public Health Surveill* 2016;2(2):e170

URL: <http://publichealth.jmir.org/2016/2/e170/>

doi: [10.2196/publichealth.6377](https://doi.org/10.2196/publichealth.6377)

PMID: [27974287](https://pubmed.ncbi.nlm.nih.gov/27974287/)

©Jessica L Maksut, Lisa A Eaton, Elizabeth J Siembida, Daniel D Driffin, Robert Baldwin. Originally published in *JMIR Public Health and Surveillance* (<http://publichealth.jmir.org>), 14.12.2016. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in *JMIR Public Health and Surveillance*, is properly cited. The complete bibliographic information, a link to the original publication on <http://publichealth.jmir.org>, as well as this copyright and license information must be included.

Viewpoint

# Lessons Learned From Methodological Validation Research in E-Epidemiology

Emmanuelle Kesse-Guyot<sup>1</sup>, PhD; Karen Assmann<sup>1</sup>, PhD; Valentina Andreeva<sup>1</sup>, PhD; Katia Castetbon<sup>2,3</sup>, PhD; Caroline Méjean<sup>1</sup>, PhD; Mathilde Touvier<sup>1</sup>, PhD; Benoît Salanave<sup>3</sup>, MD; Valérie Deschamps<sup>3</sup>, PhD; Sandrine Péneau<sup>1</sup>, PhD; Léopold Fezeu<sup>1</sup>, MD, PhD; Chantal Julia<sup>1,4</sup>, MD, PhD; Benjamin Allès<sup>1</sup>, PhD; Pilar Galan<sup>1</sup>, MD, PhD; Serge Hercberg<sup>1,4</sup>, MD, PhD

<sup>1</sup>Équipe de Recherche en Épidémiologie Nutritionnelle (EREN), Centre de Recherche en Épidémiologie et Statistiques, COMUE Sorbonne Paris Cité, Inserm (U1153), Inra (U1125), Cnam, Université Paris 13, Bobigny, France

<sup>2</sup>Ecole de Santé Publique, Centre de Recherche en Épidémiologie, Biostatistiques et Recherche Clinique, Université Libre de Bruxelles, Bruxelles, Belgium

<sup>3</sup>Unité de Surveillance et d'Épidémiologie Nutritionnelle (USEN), Institut de Veille Sanitaire, Centre de Recherche en Épidémiologie et Statistiques, COMUE Sorbonne Paris Cité, Université Paris 13, Bobigny, France

<sup>4</sup>Hôpital Avicenne, Département de Santé Publique, Bobigny, France

**Corresponding Author:**

Emmanuelle Kesse-Guyot, PhD

Équipe de Recherche en Épidémiologie Nutritionnelle (EREN)

Centre de Recherche en Épidémiologie et Statistiques

COMUE Sorbonne Paris Cité

Inserm (U1153), Inra (U1125), Cnam

Université Paris 13

74 rue Marcel Cachin

Bobigny, 93017

France

Phone: 33 0148388979

Fax: 33 0148388931

Email: [e.kesse@eren.smbh.univ-paris13.fr](mailto:e.kesse@eren.smbh.univ-paris13.fr)

**Related Article:**

This is a corrected version. See correction statement: <http://publichealth.jmir.org/2016/2/e163/>

## Abstract

**Background:** Traditional epidemiological research methods exhibit limitations leading to high logistics, human, and financial burden. The continued development of innovative digital tools has the potential to overcome many of the existing methodological issues. Nonetheless, Web-based studies remain relatively uncommon, partly due to persistent concerns about validity and generalizability.

**Objective:** The objective of this viewpoint is to summarize findings from methodological studies carried out in the NutriNet-Santé study, a French Web-based cohort study.

**Methods:** On the basis of the previous findings from the NutriNet-Santé e-cohort (>150,000 participants are currently included), we synthesized e-epidemiological knowledge on sample representativeness, advantageous recruitment strategies, and data quality.

**Results:** Overall, the reported findings support the usefulness of Web-based studies in overcoming common methodological deficiencies in epidemiological research, in particular with regard to data quality (eg, the concordance for body mass index [BMI] classification was 93%), reduced social desirability bias, and access to a wide range of participant profiles, including the hard-to-reach subgroups such as young (12.30% [15,118/122,912], <25 years) and old people (6.60% [8112/122,912], ≥65 years), unemployed or homemaker (12.60% [15,487/122,912]), and low educated (38.50% [47,312/122,912]) people. However, some selection bias remained (78.00% (95,871/122,912) of the participants were women, and 61.50% (75,590/122,912) had postsecondary education), which is an inherent aspect of cohort study inclusion; other specific types of bias may also have occurred.

**Conclusions:** Given the rapidly growing access to the Internet across social strata, the recruitment of participants with diverse socioeconomic profiles and health risk exposures was highly feasible. Continued efforts concerning the identification of specific biases in e-cohorts and the collection of comprehensive and valid data are still needed. This summary of methodological findings from the NutriNet-Santé cohort may help researchers in the development of the next generation of high-quality Web-based epidemiological studies.

(*JMIR Public Health Surveill* 2016;2(2):e160) doi:[10.2196/publichealth.5880](https://doi.org/10.2196/publichealth.5880)

## KEYWORDS

cohort studies; bias, epidemiology

## Introduction

Advances in knowledge on health and disease strongly rely on the availability, quality, and comprehensiveness of data from prospective cohort studies with very large population-based samples. Such studies are particularly important for the exploration of research hypotheses that do not lend themselves to randomized trials.

Traditional methods of data collection in large epidemiological studies (paper-and-pencil questionnaires, face-to-face interviews, and telephonic interviews) require immense resources in terms of logistics (mailing, preparation and postage, transportation, etc); personnel (interview staff, administrative help, data entry, and data quality control); and resources (printing paper, heavy-duty printers, toner cartridges, etc). In addition, traditional methods likely make study participation burdensome (eg, mailing back questionnaires, organizing appointments, transportation time to get to examination centers). As traditional epidemiological studies have been witnessing a steady decline in response rates over the last few decades [1] that has resulted in concerns about selection bias and representativeness, alternative strategies for the recruitment and follow-up of participants are gaining momentum. Although self-selection bias does not systematically lead to erroneous conclusions in etiologic research, it has major implications for investigations focusing on social determinants and for descriptive and prevalence studies [2]. Furthermore, limitations of traditional methods include interviewer-related biases and a high susceptibility to errors during data entry.

Although both traditional and e-epidemiological research relies on volunteers, the use of innovative, computerized tools could provide highly pertinent future directions [3-7]—especially in France, where 82% of the population has an access to the Internet [8].

Currently, most existing e-cohorts include young women as target populations, and are characterized by a high proportion of well-educated volunteers [9-14]. Thus far, few studies have evaluated e-cohorts with respect to participant selection processes or sociodemographic profiles, or have validated tools for Web-based epidemiological research [15-17].

In this context, the aim of this work was to compile methodological research findings from validation studies based on the ongoing NutriNet-Santé e-cohort with >100,000 participants enrolled till date.

## Methods

### General Description of the NutriNet-Santé Study

The NutriNet-Santé study [18] is a Web-based cohort launched in France in 2009, with a planned 10-year follow-up. Its main objective was the comprehensive investigation of the relationship between multiple aspects of nutrition and health [18]. Participants were recruited through a combination of Web-based and traditional recruitment strategies such as television and radio broadcasts, newspaper advertising, and flyer distribution. A secured website was used as a platform for study inclusion, data collection, and follow-up [19]. Inclusion criteria were residence in France, age  $\geq 18$  years, and access to the Internet. Participants were followed via the Internet and were asked to complete Web-based questionnaires on a regular basis. The study was approved by the ethics committee of the French Institute for Health and Medical Research (IRB INSERM n° 0000388 FWA00005831) and by the National Commission on Informatics and Liberty (CNIL n°908450 and n° 909216). All subjects signed an electronic informed consent. All data collection instruments were Web-based and provide the participants with general instructions as well as automated alerts (text, pictures, and error messages) designed to improve the accuracy and completeness of the responses.

### Sociodemographic Data

Sociodemographic data (age, gender, education, marital status, number of children, geographical region, and occupational status) were collected via Web-based questionnaires.

Participants were also asked to complete questionnaires inquiring about motives for participation (“Would you have participated to the NutriNet-Santé study if it was not Internet-based?” [yes or no]), computer and Internet skills, and channels of recruitment. They were also asked to provide part of their 15-digit national identification number (personal Social Security number).

A study was conducted to compare the self-administered Web-based version of the sociodemographic questionnaire with a traditional paper version among 147 participants (paper first, n=76; Web-based first, n=71).

### Dietary Data

Dietary data were collected at inclusion and on a biannual basis via a set of 3 24-h records (24 h) and randomly allocated over a 2-week period, including 2 weekdays and 1 weekend day. Participants reported all foods and beverages (type and quantity) consumed at each meal (breakfast, lunch, and dinner) or any



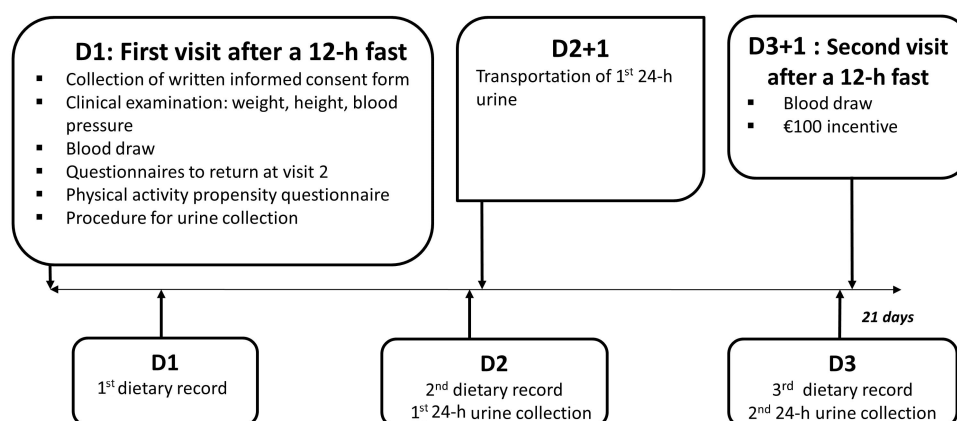
other eating occasion during a 24-h period. Each food or beverage item consumed was entered into the system via a food browser or a search engine. Additional information about time and place of eating were also collected. Portion sizes were estimated on the basis of the estimated weight of a food item, purchase units or of photographs from a validated picture booklet [20], reflecting more than 250 generic foods corresponding to more than 2000 specific food items, presented in 7 different portion sizes. Nutrient intakes were calculated using the ad hoc NutriNet-Santé composition table including more than 2000 foods and beverages [21]. In addition to the 24-h dietary records, a previously validated semiquantitative food-frequency questionnaire was proposed to the NutriNet-Santé participants 6 months after study inclusion [22].

A validation study was conducted to evaluate the quality of the data collected via the 24-h dietary record tool against 24-h

urinary and plasma biomarkers [23,24]. A total of 199 adult volunteers were included, who completed 3 24-h dietary records and provided 2 24-h urine samples and 2 blood samples concomitant with the first and third 24-h dietary record. Figure 1 illustrates the design of this validation study in a detailed manner.

Beyond the objective validity of the 24-h tool, we also carried out a comparative study to evaluate the concordance of results obtained by the 24-h Web tool with results obtained via telephonic interviews conducted by a dietician (which is the reference method used when a traditional mode of survey is administered) [25]. This study also included a comparison of the estimated financial cost related to the implementation and use of the Web-based 24-h dietary records with the cost related to the standard assessment involving dietitians.

**Figure 1.** Design of the dietary data validation study, NutriNet-Santé, 2013 (N=199).



### Anthropometric Data

Anthropometric data including weight, height, restrictive dieting, and weight history were collected through a self-administered questionnaire at inclusion and annually thereafter. In addition, a subsample of participants (n=20,000) completed clinical evaluations and biological sampling during clinic visits held throughout France. During these visits, participants underwent a clinical examination including weight, height, waist circumference, hip circumference, and bioimpedance measurements. In order to evaluate the quality of the self-reported anthropometric data, a random subsample of all participants with a scheduled clinical examination were asked to complete a Web-based anthropometric questionnaire 3 days before their appointment (November 2011-July 2012) [26]. Among this subsample of participants, a further subsample was randomly selected and asked to verbally report their height and weight to a trained technician on the day of the examination before being measured.

An additional comparative study was set-up in order to compare the Web-based anthropometric questionnaire with a paper-based version of the same questionnaire [27].

### Statistical Analyses Performed in the Different Validation Studies

In order to provide a detailed presentation of the sociodemographic profiles of the study population, characteristics of 122,912 NutriNet-Santé participants (recruited between 2009 and 2014) were compared with data from the French 2009 Census [28]. Moreover, the impact of a statistical weighting method developed to enhance the representativeness of descriptive or analytical results was evaluated, using the SAS macro "CALMAR (CALage sur MARGes)" provided by the Institut national de la statistique et des études économiques (INSEE) [29]. This macro was developed to recover a population-representative sample from nonrepresentative samples obtained by surveys by weighting individual data using ancillary information. In the study on demographic data in the NutriNet-Santé cohort, weights were calculated based on gender, age, birthplace, educational level, occupation, marital status, presence of children in the household, and area of residence.

Dietary data in the NutriNet-Santé cohort were compared with nationally representative data from the cross-sectional Etude Nationale Nutrition Santé (ENNS, 2006-2007), including 2754 French adults aged 18-74 years [30].

In both studies, reported dietary intakes from 3 (nonconsecutive days over a 2-week period, including 2 weekdays and 1 weekend day) 24-h dietary surveys (computerized records in the



NutriNet-Santé study and telephone recalls in the ENNS) were weighted using the French Census estimates. The same food composition table was used to estimate nutrient intakes in both samples.

Several statistical indicators were used to analyze the data including means, standard errors of the mean, intraclass correlation coefficients (ICCs), kappa coefficients, and Spearman correlation coefficients. Self-reported dietary intakes of protein, potassium, and sodium were compared with intakes estimated on the basis of urinary biomarkers, using a validated log-ratio formula [23]. Both simple Spearman correlation coefficients and coefficients adjusted for age, BMI, smoking, education, energy intake, alcohol consumption, and use of dietary supplements were calculated.

## Results

### Summary

Our findings are based on 12 studies investigating characteristics and representativeness of the sample [28], comparison of dietary data with national findings [31], recruitment [32], motives for participation [33], Internet or computer skills [34], and quality of the data [23-27,35,36].

### Characteristics and Representativeness

The gender ratio (females:males) in the e-cohort was 5:3 and the average age of participants at inclusion was 42.6 (SD 14.6) years. A small proportion, that is, 5.00% (6208/122,912) of participants were born outside of France, 31.40% (38,606/122,912) had university-level education, 70.80% (87,048/

122,912) were married or cohabiting, and 65.80% (80,934/122,912) did not have children aged under 18 years in their household. [Figure 2](#) presents the crude (unweighted) sociodemographic characteristics of the NutriNet-Santé study population in comparison with the French Census data from 2009 [28]. The NutriNet-Santé sample was relatively similar to the national data concerning the geographical area of residence and the presence of children in the household.

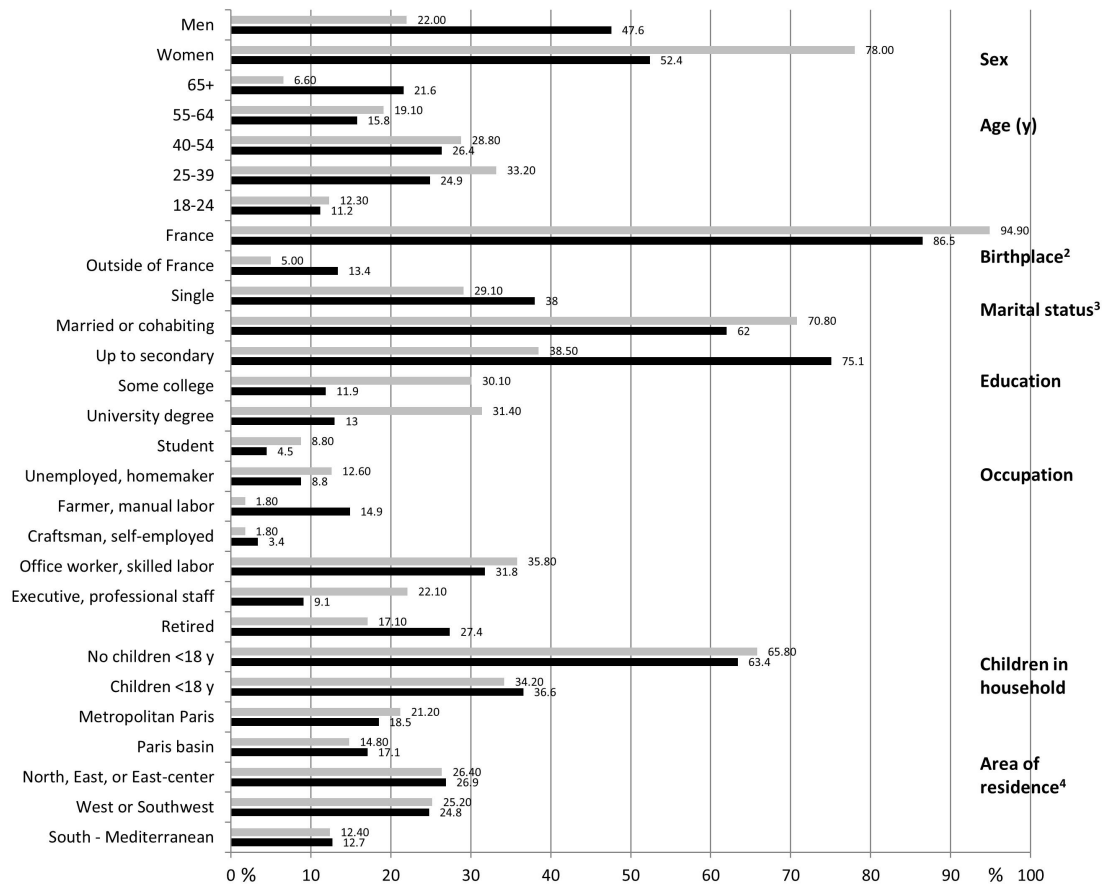
Similarly, the age distribution up to 65 years resembled the respective data observed in the general French population. However, important discrepancies were observed concerning gender, education, and occupation. Some subgroups are highly underrepresented in the e-cohort (eg, farmers), limiting the effectiveness of the weighting method and thus the generalizability. It is noteworthy that the proportion of volunteers who were unemployed or homemakers or disabled was higher in the NutriNet-Santé sample than in the general French population.

### Comparison of Dietary Data With National Findings

A comparison of dietary intake in the NutriNet-Santé and the nationally representative ENNS is shown in [Figure 3](#) and [Table 1](#) [31].

The comparison revealed that the consumption of fruits and vegetables and fish and seafood was higher in the NutriNet-Santé cohort. In contrast, the consumption of potatoes and pulses, meat and eggs, and nonalcoholic beverages was significantly lower in the NutriNet-Santé cohort than in the ENNS.

**Figure 2.** Comparison of the sociodemographic characteristics of NutriNet-Santé (2009-2014) subjects (N=122,912) with French census data. Gray bars denote NutriNet-Santé subjects. Black bars denote French Census estimates (2009) for individuals aged 18 years and above in metropolitan France from INSEE. All differences between NutriNet-Santé subjects and the Census estimates were statistically significant (all chi-square-test *P* values were <.001).<sup>2</sup>France: including Corsica and overseas territories.<sup>3</sup>Single: never-married, widowed, divorced, or separated.<sup>4</sup>Geographical distribution based on the Zone d'études et d'aménagement du territoire (ZEAT) defined by INSEE. INSEE: Institut national de la statistique et des études économiques.



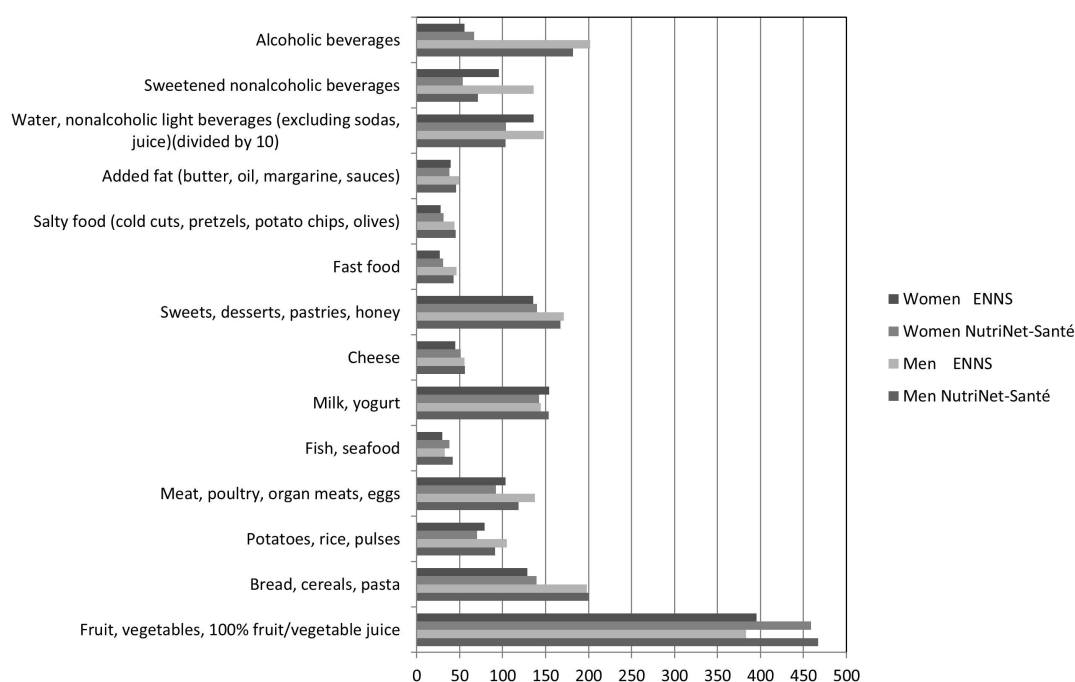
**Table 1.** Intake of nutrients in the NutriNet-Santé study (2009-2010, N=49,443) and the nationally representative survey (ENNS, 2006-2007, n=2754)<sup>a,b</sup>.

Nutritional indicators	Men		Women	
	NutriNet-Santé	ENNS <sup>c</sup>	NutriNet-Santé	ENNS
Total energy (Kcal/d)	2326.31 (5.79)	2388.67 (27.70)	1767.94 (2.21)	1713.69 (14.02)
Total carbohydrates (g/d)	238.61 (0.69)	246.11 (3.33)	184.04 (0.27)	180.35 (1.69)
Total lipids (g/d)	97.11 (0.32)	98.44 (1.27)	75.95 (0.13)	73.30 (0.74)
Protein (g/d)	96.13 (0.26)	98.31 (1.13)	75.72 (0.10)	74.10 (0.70)
Dietary fiber (g/d)	21.30 (0.08)	19.12 (0.30)	18.17 (0.03)	16.10 (0.19)
Calcium (mg/d)	1028.98 (3.48)	1022.16 (13.91)	879.87 (1.50)	869.80 (9.94)
Retinol (µg/d)	608.97 (8.01)	668.23 (37.71)	480.38 (3.72)	496.11 (27.02)
Beta-carotene (µg/d)	3418.12 (27.22)	3196.52 (113.87)	3270.61 (13.40)	3211.01 (67.38)
Vitamin B6 (mg/d)	2.04 (0.01)	1.89 (0.02)	1.63 (0.00)	1.52 (0.02)
Vitamin B9 (µg/d)	352.74 (1.27)	332.61 (4.78)	312.18 (0.59)	292.16 (3.15)
Vitamin B12 (µg/d)	6.36 (0.06)	6.17 (0.21)	5.02 (0.03)	4.66 (0.14)
Vitamin C (mg/d)	117.53 (0.78)	95.87 (3.49)	109.87 (0.44)	96.68 (1.70)
Vitamin D (µg/d)	2.93 (0.02)	2.47 (0.10)	2.50 (0.01)	2.00 (0.06)
Vitamin E (mg/d)	13.16 (0.05)	11.72 (0.22)	11.10 (0.03)	9.62 (0.12)
Zinc (mg/d)	13.06 (0.04)	13.09 (0.18)	10.29 (0.02)	9.67 (0.11)
Iron (mg/d)	15.63 (0.06)	14.00 (0.20)	12.45 (0.02)	10.80 (0.12)
Potassium (mg/d)	3344.58 (9.15)	3194.87 (38.86)	2840.51 (3.99)	2668.44 (24.74)
Magnesium (mg/d)	372.75 (1.19)	329.68 (3.68)	310.03 (0.52)	266.18 (2.54)

<sup>a</sup>All data from both NutriNet-Santé and ENNS are weighted for age, education, presence of children in the household, and season of data collection, using French 2007 Census figures.

<sup>b</sup>Values are means with SEs of the means within the brackets.

<sup>c</sup>ENNS: Etude Nationale Nutrition Santé.

**Figure 3.** Mean food (g/d) and beverage (ml/d) intake in the NutriNet-Santé study (2009-2010, N=49,443) and the nationally representative survey (ENNS, 2006-2007, n=2754). All data from both NutriNet-Santé and ENNS are weighted for age, education, presence of children in household, and season of data collection, using French 2007 Census data. ENNS: Etude Nationale Nutrition Santé.

Total energy and macronutrient intake was similar overall, whereas intake of fiber, iron, magnesium, and vitamins B6, B9, C, D, and E were higher in the NutriNet-Santé study than in the ENNS. Overall, true differences in intake, differential measurement errors related to the different modes of administration, sample composition (volunteers vs stratified, random sample), and volunteer bias might each contribute to explaining the findings. Concerning potential volunteer bias in the NutriNet-Santé study, it should be noted the magnitude of such bias should be similar in volunteer-based nutritional epidemiologic studies that use traditional methods of data collection. Given the large samples in both the NutriNet-Santé study and the ENNS study, it is noteworthy that not all statistically significant differences are practically meaningful.

### **Recruitment, Motives for Participation, and Internet or Computer Skills**

Efforts for diversifying communication channels for sample recruitment could be an important element for increasing participant profile heterogeneity. We have thus investigated the different channels through which subjects were ultimately recruited according to sociodemographic profiles [32].

The largest number of subjects was reached using television advertisements (34.46%, 30,414/88,238). Other channels with substantial importance were radio broadcasts (16.21%, 14,309 out of 88,238) and websites (19.05%, 16,807/88,238). Note that compared with other channels, among subjects recruited by television broadcasts, there was a higher proportion of individuals with a low educational level (44.1% vs 37.4% in the whole sample) and of unemployed individuals (36.1% vs 4.8% in the whole sample).

As specific profiles of subjects involved in nutritional Web-based cohorts may be related to specific motives for participation, we have investigated reasons and motives for participation in the NutriNet-Santé study according to sociodemographic characteristics [33]. The use of Internet as

an exclusive means of follow-up was an important source of motivation for participation, especially among men, young individuals, those of high socioeconomic status, and among obese persons.

A substantial portion (23.79%, 10,235 out of 43,028) of subjects in the NutriNet-Santé study has relatively low self-reported computer or Internet skills, suggesting that this factor does not represent a barrier to participation in Web-based cohorts [34]. Besides, several subgroups of subjects with lower computer skills (eg, women or those with lower educational level) seemed more inclined to accept the demands associated with participation in the Web cohort [34].

### **Quality of the Collected Data**

#### *Sociodemographic Data*

Agreement between data obtained from the self-administered Web-based version and the traditional paper version of the sociodemographic questionnaire was high, with ICCs for continuous variables ranging between .81 and 1 and kappa coefficients for categorical variables ranging between .76 and 1. Administration order, age, gender, and Internet skills did not extensively modify the results [36].

Response consistency between data from the personal Social Security number and information collected through the sociodemographic questionnaire was estimated among 84,442 subjects (84.34%, 84,442/100,118). A total of 5141 subjects (6.09%, 5141/84,442) had at least one discordant data value. Sex, age, education, and employment were associated with the response consistency patterns [35].

#### *Dietary Data*

The comparison of self-reported dietary data against urinary biomarkers (Table 2) revealed that, on an average, men underreported protein intake, whereas women underreported both protein and sodium intake [23].

**Table 2.** Dietary intake and urinary excretion of protein, potassium, and sodium, NutriNet-Santé study, 2013<sup>a</sup>.

Nutrients and indicators	Men (n=102)			Women (n=91)		
	Mean	Lower value of the 95% CI	Upper value of the 95% CI	Mean	Lower value of the 95% CI	Upper value of the 95% CI
<b>Protein</b>						
Mean 24-h U <sup>b,d</sup> (g/day)	101.7	62.3	166.2	77.4	45.8	130.5
Mean 24 h <sup>b,c</sup> (g/day)	88.6	83.9	93.7	68.8	65.1	72.8
Difference % <sup>c</sup>	-14.4	-18.2	-10.3	-13.9	-18.3	-9.3
<b>Potassium</b>						
Mean 24-h U <sup>b</sup> (mg/day)	3357	3189	3535	2843	2685	3010
Mean 24 h <sup>b</sup> (mg/day)	3444	3279	3618	2739	2607	2879
Difference % <sup>c</sup>	2.6	-1.7	7.1	-3.6	-8.9	1.9
<b>Sodium</b>						
Mean 24-h U <sup>b</sup> (mg/day)	3578	3320	3856	2996	2790	3217
Mean 24 h <sup>b</sup> (mg/day)	3503	3271	3752	2747	2567	2941
Difference % <sup>c</sup>	-2.1	-9.2	5.6	-8.3	-15.7	-0.2

<sup>a</sup>Dietary intake values are mean values across 3 24 h, and urinary excretion is the mean of 2 24-h urine samples.

<sup>b</sup>Geometric means based on log-transformed data.

<sup>c</sup>Mean difference in percentage calculated from the log-ratio of mean reported intake (24 h) over mean biomarker intake (24-h Us).

<sup>d</sup>24-h U, 24-h urine collection.

<sup>e</sup>24-h, 24-h dietary record.

Relative differences between reported and “measured” intakes were -14.4% (protein), 2.6% (potassium), -2.1% (sodium) for men, and -13.9%, -3.6%, and -8.3% for women. Misreporting was not associated with body weight status.

Among men, the investigation revealed Spearman correlations ranging between .20 (for vegetables and plasma vitamin C) and .55 (for fish and plasma docosahexaenoic acid [DHA]). Among women, these correlations were generally lower and ranged from .13 (nonsignificant; for vegetables and plasma vitamin C) to .54 (for fish and eicosapentaenoic acid [EPA]+DHA). Regarding micronutrients, adjusted correlations of self-reported intakes with plasma biomarkers ranged from .36 (EPA) to .58 (vitamin C) in men and from .32 (vitamin C) to .38 (EPA) in women [24].

Next, the agreement between the 24-h Web tool data and data obtained via telephone interviews was very high [37], with ICCs ranging from .5 for fats or sauces (among both genders), breakfast cereals, cakes or biscuits or pastries and dairy (women

only to .9 for fruits, pulses (among both genders), breakfast cereals, alcoholic drinks, and meat (among men). For nutrient intake, energy-adjusted Pearson correlation coefficients ranged from .6 for polyunsaturated fatty acid, retinol, vitamin E, and sodium (among women) to .9 for a large number of different nutrients. Note that women participating in this comparative study reported higher intake of cakes, biscuits, or pastries when using the Web interface than during the telephone interview with a dietician, indicating that social desirability bias may have been lower in the Web-based version as compared with the traditional version. The acceptability of the Web-based nutrition assessment tool was high, with 92.74% (115/124) of participants judging the Web interface as user-friendly and 66.13% (82/124) preferring the Web-based method over the interview.

The estimation of the financial cost related to the implementation and use of the Web-based dietary questionnaire compared with the cost associated with the standard assessment involving dietitians (for 1 24-h dietary record/recall) are presented in Table 3.



**Table 3.** Cost estimation for traditional and Web-based assessment<sup>a</sup>.

Questionnaire	Cost
<b>Traditional methods for baseline data assessment</b>	
Dietary data, interview (for 1 24-h recall) comprising the salary of the dietitians, telephone expenses, and cost of printing and sending the picture booklet	€38.1/subject
Anthropometric data, paper version comprising printing, postage for sending and returning the questionnaire, and double data entry	€9.9/subject
Sociodemographic data, paper version comprising printing, postage for sending and returning the questionnaire, and double data entry	€16.5/subject
Total for 100,000 subjects	€6,450,000
<b>Web-based method</b>	
NutriNet-Santé Web-based platform comprising the whole study process: secure registration system, development, and administration of baseline questionnaires (including three 24-h records), license, equipment, and hosting)	€380,000
No supplementary cost for an additional dietary assessment	

<sup>a</sup>Financial estimation using cost in 2009.

**Table 4.** Advantages of Web-based cohort studies for data collection with respect to traditional modes of epidemiological research.

Problematic aspects encountered in epidemiological studies	Advantages of Web-based cohort studies
Representativeness	The use of Internet as the exclusive mode of follow-up was a decisive reason for participation, in particular for men, young individuals, and obese persons Television broadcasts may help to further increase the proportion of population groups that are less likely to participate in cohorts of volunteers (young people, elderly, men, and low socioeconomic status groups) Individuals with a lower computer literacy level might participate Individuals with hearing and other disabilities might participate
Data collection	Reduced logistic, personnel, material, and financial burden of large epidemiological studies Reduction of data entry errors as compared with paper questionnaires
Acceptability	Data entry can be rendered easier and more pleasant by the use of well-designed interactive interfaces and videos, unlike paper questionnaires
Data management	Data treatment can be directly incorporated into the software, leading to rapid availability of the collected information
Validity of nutritional and anthropometric data	High or similar quality as in conventional studies Concerning nutritional data, social desirability bias may be lower in Web-based studies (in particular among women)
Development of new assessment tools	Possibility to rapidly test and implement new assessment tools, protocols, and so on.
Health events assessments	Possibility to rapidly match participant data with different medical registries

### **Anthropometric Data**

The validation study showed high ICCs, ranging from .94 for height to .99 for weight, and the concordance for BMI classification was 93% (sensitivity 88% and specificity 99%). However, we observed a slight underreporting of weight and overreporting of height, leading to an underreporting of BMI, which was more pronounced among obese participants. Web-based and face-to-face self-reports of weight and height were almost perfectly concordant (classification agreement was 98.5%).

In the comparative study, agreement between the Web-based version of the self-administered anthropometric questionnaire and the paper-based version of the same questionnaire was very high, with ICCs ranging from .86 to 1 and kappa statistics

ranging from .69 to 1 for continuous and categorical variables, respectively [27].

Overall, based on the results of the methodological studies carried out in the NutriNet-Santé study, we provide new insights with respect to Internet use in epidemiological research (Table 4).

## **Discussion**

### **Principal Findings**

Overall, the presented findings support an acceptable to high quality of data collected in a large and heterogeneous Web-based cohort with a substantially reduced financial burden. Concordance assessed in objective validation studies and in method comparison studies was high. Moreover, the presented

results indicate that Web-based studies may help reduce social desirability biases.

It appears that the validity of the collected data was comparable with that in conventional research (based on interviews or paper-and-pencil tools), or even higher. For instance, acceptable to moderate validity of nutrient intake estimation was observed. Although the observed correlation coefficients may be considered as relatively low in absolute terms, previous validation studies have reported correlations of a similar magnitude [38-41]. Next, bias associated with social desirability, inherent in studies based on interviews, is likely lower in Web-based studies due to higher perceived anonymity particularly among women [6].

Cohorts of volunteers like the NutriNet-Santé study tend to have a distribution of sociodemographic profiles that diverges from the general population [1]. In particular, prospective nutritional cohorts relying on volunteer-based samples have been criticized with respect to their strong susceptibility to self-selection bias. It has been shown that women, older individuals, and married individuals were more prone to enroll in epidemiological studies, irrespective of the research topic, whereas at-risk populations were less likely to participate [1,42,43]. In addition, it has been postulated that Web-based cohorts including motivated Internet-skilled volunteers may be particularly at risk for such bias as Internet accessibility and patterns of Internet use may vary according to sociodemographic profile [7]. However, in Europe, this additional source of selection bias is likely minor given the wide access to the Internet in all subgroups of the population including people above 55 years (42% in 2011) and those with low education (45% in 2011) [44]. Yet, it is possible that difficulties with Internet use may have caused a number of participants with low Internet skills or Internet access problems, in particular older people, to drop out of the study. Beyond self-reported computer skills, the type of activity (unexpected clicks, nonobservance of instructions, etc) across subgroups also merits evaluation.

Our findings also suggest that the exclusive use of Internet for data collection and follow-up—which implies a largely reduced effort for participants [3,45] and research staff—may help increase the relative proportion of population subgroups (young people, elderly, men) that are usually underrepresented [1,42,43]. Nevertheless, the Web-based design appeared to be a less important motive for participation among individuals in “low” compared with “high” occupational categories. Besides, the use of a wide range of recruitment channels beyond the Internet may help further diversify exposure profiles. Television, which is a wide-reaching medium in France, may be a particularly useful instrument for improving the recruitment of population subgroups that tend to be underrepresented in volunteer-based epidemiologic studies such as low educated individuals or unemployed individuals.

### Advantages and Limitations

Web-based cohorts have the potential to include a large number of participants, including the hard-to-reach subgroups. Yet, as those who are willing to participate in a Web-based cohort study probably differ from the general population, other means to

specifically increase opportunities for participation in epidemiological Web-based studies among low socioeconomic groups is needed. In that context, statistical weighting methods are an important and efficient approach to help counterbalance limitations concerning the external validity of descriptive epidemiological studies.

A key advantage of Web-based research is related to data entry. By comparing the Web-based version of baseline questionnaires with the traditional paper-and-pencil version used in methodological studies [27,36], we also showed that a substantial number of data entry mistakes, missing values, and inconsistent or abnormal values were found in the paper version, whereas they were nonexistent in the Web-based version due to integrated controls. Computerized data entry also obviates the need for data coding or entry personnel. This provides major advantages concerning the logistic, personnel, material, and financial burden of large epidemiologic studies. In addition, a large proportion of data quality control can directly be implemented into the Web-based software that may substantially improve the quality of the collected data [6]. In contrast to a paper questionnaire, Web-based tools can directly alert the participant if unrealistic values are entered or prevent the participant from submitting an incomplete questionnaire. Besides, long questionnaires can be simplified by implementing conditional jumps.

In addition, data entry can be rendered more pleasant by the use of a well-designed interactive interface, audio and visual feedback, and pop-up windows providing additional information regarding potentially complicated elements of the questionnaires. Indeed, our investigation of the quality of the nutritional data collected in NutriNet-Santé showed that a very high percentage of participants perceived the Web-based nutrition assessment tool as user-friendly. A further advantage pertains to the high flexibility of the Web-based platform, allowing rapid implementation of new questionnaires or ancillary protocols. For instance, several questionnaires have been developed or translated in the context of the NutriNet-Santé study, as such a platform permits rapid testing and fine-tuning for the validation of new questionnaires [46,47].

### Future Research

This paper provides an overview of methodological advances in e-epidemiology and fills gaps in knowledge concerning specific methodological aspects of e-cohorts and their design, such as participant selection processes, sample representativeness, diversity of sociodemographic profiles, and the validation of Web-based tools. The development of ad hoc validation and comparative studies may help improve innovative digital tools and to reinforce confidence in data collected using new technologies. This may guide further development and implementation of future e-cohorts and validation studies that are currently scant.

Future directions may include evaluating the consistency of health data collected via different sources (self-reported data, disease registries, or medical databases). It would also be useful to implement ad hoc studies to compare characteristics of hard-to-reach subgroups with those of nonparticipants.

## Acknowledgments

The NutriNet-Santé study is supported by the French Ministry of Health (DGS), the French Institute for Health Surveillance (InVS), the National Institute for Prevention and Health Education (INPES), the Foundation for Medical Research (FRM), the National Institute for Health and Medical Research (INSERM), the National Institute for Agricultural Research (INRA), the National Conservatory of Arts and Crafts (CNAM), and the University of Paris 13. The ENNS study was funded by the French Institute for Health Surveillance (InVS), the French National Health Insurance System (Caisse Nationale d'Assurance Maladie des Travailleurs Salariés, CnamTS) and the University of Paris 13. The authors express their gratitude to the data managers and statisticians involved with the NutriNet-Santé study, and to the dieticians who collected ENNS data.

## Conflicts of Interest

None declared.

## References

1. Galea S, Tracy M. Participation rates in epidemiologic studies. *Ann Epidemiol* 2007 Sep;17(9):643-653. [doi: [10.1016/j.annepidem.2007.03.013](https://doi.org/10.1016/j.annepidem.2007.03.013)] [Medline: [17553702](https://pubmed.ncbi.nlm.nih.gov/17553702/)]
2. Nohr EA, Frydenberg M, Henriksen TB, Olsen J. Does low participation in cohort studies induce bias? *Epidemiology* 2006 Jul;17(4):413-418. [doi: [10.1097/01.ede.0000220549.14177.60](https://doi.org/10.1097/01.ede.0000220549.14177.60)] [Medline: [16755269](https://pubmed.ncbi.nlm.nih.gov/16755269/)]
3. Schatzkin A, Subar AF, Moore S, Park Y, Potischman N, Thompson FE, et al. Observational epidemiologic studies of nutrition and cancer: the next generation (with better observation). *Cancer Epidemiol Biomarkers Prev* 2009 Apr;18(4):1026-1032 [FREE Full text] [doi: [10.1158/1055-9965.EPI-08-1129](https://doi.org/10.1158/1055-9965.EPI-08-1129)] [Medline: [19336550](https://pubmed.ncbi.nlm.nih.gov/19336550/)]
4. Ekman A, Dickman PW, Klint A, Weiderpass E, Litton JE. Feasibility of using web-based questionnaires in large population-based epidemiological studies. *Eur J Epidemiol* 2006;21(2):103-111. [doi: [10.1007/s10654-005-6030-4](https://doi.org/10.1007/s10654-005-6030-4)] [Medline: [16518678](https://pubmed.ncbi.nlm.nih.gov/16518678/)]
5. Ekman A, Klint A, Dickman PW, Adami HO, Litton JE. Optimizing the design of web-based questionnaires--experience from a population-based study among 50,000 women. *Eur J Epidemiol* 2007;22(5):293-300. [doi: [10.1007/s10654-006-9091-0](https://doi.org/10.1007/s10654-006-9091-0)] [Medline: [17206467](https://pubmed.ncbi.nlm.nih.gov/17206467/)]
6. van Gelder MM, Bretveld RW, Roeleveld N. Web-based questionnaires: the future in epidemiology? *Am J Epidemiol* 2010 Dec 1;172(11):1292-1298 [FREE Full text] [doi: [10.1093/aje/kwq291](https://doi.org/10.1093/aje/kwq291)] [Medline: [20880962](https://pubmed.ncbi.nlm.nih.gov/20880962/)]
7. Ekman A, Litton JE. New times, new needs; e-epidemiology. *Eur J Epidemiol* 2007;22(5):285-292. [doi: [10.1007/s10654-007-9119-0](https://doi.org/10.1007/s10654-007-9119-0)] [Medline: [17505896](https://pubmed.ncbi.nlm.nih.gov/17505896/)]
8. Bigot R, Croutte P. La diffusion des technologies de l'information et de la communication dans la société française. 317. CREDOC. Collection des rapports 2014.
9. Russell CW, Boggs DA, Palmer JR, Rosenberg L. Use of a web-based questionnaire in the Black Women's Health Study. *Am J Epidemiol* 2010 Dec 1;172(11):1286-1291 [FREE Full text] [doi: [10.1093/aje/kwq310](https://doi.org/10.1093/aje/kwq310)] [Medline: [20937635](https://pubmed.ncbi.nlm.nih.gov/20937635/)]
10. Smith B, Smith TC, Gray GC, Ryan MA, Millennium Cohort Study Team. When epidemiology meets the Internet: Web-based surveys in the Millennium Cohort Study. *Am J Epidemiol* 2007 Dec 1;166(11):1345-1354 [FREE Full text] [doi: [10.1093/aje/kwm212](https://doi.org/10.1093/aje/kwm212)] [Medline: [17728269](https://pubmed.ncbi.nlm.nih.gov/17728269/)]
11. West R, Gilsean A, Coste F, Zhou X, Brouard R, Nonnemaker J, et al. The ATTEMPT cohort: a multi-national longitudinal study of predictors, patterns and consequences of smoking cessation; introduction and evaluation of internet recruitment and data collection methods. *Addiction* 2006 Sep;101(9):1352-1361. [doi: [10.1111/j.1360-0443.2006.01534.x](https://doi.org/10.1111/j.1360-0443.2006.01534.x)] [Medline: [16911736](https://pubmed.ncbi.nlm.nih.gov/16911736/)]
12. Turner C, Bain C, Schluter PJ, Yorkston E, Bogossian F, McClure R, Nurses and Midwives e-cohort Group. Cohort Profile: The Nurses and Midwives e-Cohort Study--a novel electronic longitudinal study. *Int J Epidemiol* 2009 Feb;38(1):53-60 [FREE Full text] [doi: [10.1093/ije/dym294](https://doi.org/10.1093/ije/dym294)] [Medline: [18202083](https://pubmed.ncbi.nlm.nih.gov/18202083/)]
13. Richiardi L, Baussano I, Vizzini L, Douwes J, Pearce N, Merletti F. Feasibility of recruiting a birth cohort through the Internet: the experience of the NINFEA cohort. *Eur J Epidemiol* 2007;22(12):831-837. [doi: [10.1007/s10654-007-9194-2](https://doi.org/10.1007/s10654-007-9194-2)] [Medline: [17955333](https://pubmed.ncbi.nlm.nih.gov/17955333/)]
14. Mikkelsen EM, Hatch EE, Wise LA, Rothman KJ, Riis A, Sørensen HT. Cohort profile: the Danish Web-based Pregnancy Planning Study--'Snart-Gravid'. *Int J Epidemiol* 2009 Aug;38(4):938-943 [FREE Full text] [doi: [10.1093/ije/dyn191](https://doi.org/10.1093/ije/dyn191)] [Medline: [18782897](https://pubmed.ncbi.nlm.nih.gov/18782897/)]
15. Pizzi C, De Stavola BL, Pearce N, Lazzarato F, Ghiotti P, Merletti F, et al. Selection bias and patterns of confounding in cohort studies: the case of the NINFEA web-based birth cohort. *J Epidemiol Community Health* 2012 Nov;66(11):976-981. [doi: [10.1136/jech-2011-200065](https://doi.org/10.1136/jech-2011-200065)] [Medline: [22147750](https://pubmed.ncbi.nlm.nih.gov/22147750/)]
16. Nilsen RM, Suren P, Gunnes N, Alsaker ER, Bresnahan M, Hirtz D, et al. Analysis of self-selection bias in a population-based cohort study of autism spectrum disorders. *Paediatr Perinat Epidemiol* 2013;27(6):553-563. [Medline: [23919580](https://pubmed.ncbi.nlm.nih.gov/23919580/)]

17. Pugh CA, Summers KM, Bronsvoort BM, Handel IG, Clements DN. Validity of Internet-based longitudinal study data: the elephant in the virtual room. *J Med Internet Res* 2015;17(4):e96 [FREE Full text] [doi: [10.2196/jmir.3530](https://doi.org/10.2196/jmir.3530)] [Medline: [25887101](https://pubmed.ncbi.nlm.nih.gov/25887101/)]
18. Hercberg S, Castetbon K, Czernichow S, Malon A, Mejean C, Kesse E, et al. The Nutrinet-Santé Study: a web-based prospective study on the relationship between nutrition and health and determinants of dietary patterns and nutritional status. *BMC Public Health* 2010;10:242 [FREE Full text] [doi: [10.1186/1471-2458-10-242](https://doi.org/10.1186/1471-2458-10-242)] [Medline: [20459807](https://pubmed.ncbi.nlm.nih.gov/20459807/)]
19. Etude-nutrinet-sante. Welcome to Study the Nutrient Health [in French] URL: <http://www.etude-nutrinet-sante.fr/> [accessed 2016-10-12] [WebCite Cache ID 6lCRhHEm5]
20. Le MN, Deheeger M, Preziosi P, Montero P, Valeix P, Rolland-Cachera M, et al. Validation du manuel photos utilisé pour l'enquête alimentaire de l'étude. *SU.VI.MAX. Cahier de Nutrition et de Diététique* 1996;31(1):158-164.
21. NutriNet-Santé C. Table de composition des aliments - Etude NutriNet-Santé. In: *Economica*. Paris: Inserm; 2013.
22. Kesse-Guyot E, Castetbon K, Touvier M, Hercberg S, Galan P. Relative validity and reproducibility of a food frequency questionnaire designed for French adults. *Ann Nutr Metab* 2010;57(3):153-162. [Medline: [21079389](https://pubmed.ncbi.nlm.nih.gov/21079389/)]
23. Lassale C, Castetbon K, Laporte F, Camilleri GM, Deschamps V, Vernay M, et al. Validation of a Web-based, self-administered, non-consecutive-day dietary record tool against urinary biomarkers. *Br J Nutr* 2015;113(6):953-962. [Medline: [25772032](https://pubmed.ncbi.nlm.nih.gov/25772032/)]
24. Lassale C, Castetbon K, Laporte F, Deschamps V, Vernay M, Camilleri GM, et al. Correlations between Fruit, Vegetables, Fish, Vitamins, and Fatty Acids Estimated by Web-Based Nonconsecutive Dietary Records and Respective Biomarkers of Nutritional Status. *J Acad Nutr Diet* 2016 Mar;116(3):427-438. [doi: [10.1016/j.jand.2015.09.017](https://doi.org/10.1016/j.jand.2015.09.017)] [Medline: [26522988](https://pubmed.ncbi.nlm.nih.gov/26522988/)]
25. Touvier M, Kesse-Guyot E, Méjean C, Pollet C, Malon A, Castetbon K, et al. Comparison between an interactive web-based self-administered 24 h dietary record and an interview by a dietitian for large-scale epidemiological studies. *Br J Nutr* 2011 Apr;105(7):1055-1064. [doi: [10.1017/S0007114510004617](https://doi.org/10.1017/S0007114510004617)] [Medline: [21080983](https://pubmed.ncbi.nlm.nih.gov/21080983/)]
26. Lassale C, Péneau S, Touvier M, Julia C, Galan P, Hercberg S, et al. Validity of web-based self-reported weight and height: results of the Nutrinet-Santé study. *J Med Internet Res* 2013;15(8):e152 [FREE Full text] [doi: [10.2196/jmir.2575](https://doi.org/10.2196/jmir.2575)] [Medline: [23928492](https://pubmed.ncbi.nlm.nih.gov/23928492/)]
27. Touvier M, Mejean C, Kesse-Guyot E, Pollet C, Malon A, Castetbon K, et al. Comparison between web-based and paper versions of a self-administered anthropometric questionnaire. *Eur J Epidemiol*; 2010;25(5):287-296.
28. Andreeva V, Salanave B, Castetbon K, Deschamps V, Vernay M, Kesse-Guyot E, et al. Comparison of the sociodemographic characteristics of the large NutriNet-Santé e-cohort with French Census data: the issue of volunteer bias revisited. *J Epidemiol Community Health* 2015 Sep;69(9):893-898. [doi: [10.1136/jech-2014-205263](https://doi.org/10.1136/jech-2014-205263)] [Medline: [25832451](https://pubmed.ncbi.nlm.nih.gov/25832451/)]
29. Insee. La macro SAS CALMAR [In French] URL: [http://www.insee.fr/fr/methodes/default.asp?page=outils/calmar/accueil\\_calmar.htm](http://www.insee.fr/fr/methodes/default.asp?page=outils/calmar/accueil_calmar.htm) [accessed 2016-07-26] [WebCite Cache ID 6jI9N91mT]
30. Castetbon K, Vernay M, Malon A, Salanave B, Deschamps V, Roudier C, et al. Dietary intake, physical activity and nutritional status in adults: the French nutrition and health survey (ENNS, 2006-2007). *Br J Nutr* 2009 Sep;102(5):733-743. [doi: [10.1017/S0007114509274745](https://doi.org/10.1017/S0007114509274745)] [Medline: [19250574](https://pubmed.ncbi.nlm.nih.gov/19250574/)]
31. Andreeva V, Deschamps V, Salanave B, Castetbon K, Verdout C, Kesse-Guyot E, et al. Dietary intake comparison between a large online cohort (Etude NutriNet-Santé) and a nationally-representative cross-sectional study in France (ENNS): addressing the issue of generalizability in e-epidemiology. *Am J Epidemiol* 2016 (forthcoming)(forthcoming).
32. Kesse-Guyot E, Andreeva V, Castetbon K, Vernay M, Touvier M, Mejean C, et al. Participant profiles according to recruitment source in a large Web-based prospective study: experience from the Nutrinet-Sante study. *J Med Internet Res* 2013;15(9):e205. [Medline: [24036068](https://pubmed.ncbi.nlm.nih.gov/24036068/)]
33. Méjean C, Szabo de Edelenyi F, Touvier M, Kesse-Guyot E, Julia C, Andreeva VA, et al. Motives for participating in a web-based nutrition cohort according to sociodemographic, lifestyle, and health characteristics: the NutriNet-Santé cohort study. *J Med Internet Res* 2014;16(8):e189 [FREE Full text] [doi: [10.2196/jmir.3161](https://doi.org/10.2196/jmir.3161)] [Medline: [25135800](https://pubmed.ncbi.nlm.nih.gov/25135800/)]
34. Pouchieu C, Mejean C, Andreeva VA, Kesse-Guyot E, Fassier P, Galan P, et al. How computer literacy and socioeconomic status affect attitudes toward a Web-based cohort: results from the NutriNet-Sante study. *J Med Internet Res* 2015;17(2):e34. [Medline: [25648178](https://pubmed.ncbi.nlm.nih.gov/25648178/)]
35. Andreeva VA, Galan P, Julia C, Castetbon K, Kesse-Guyot E, Hercberg S. Assessment of response consistency and respective participant profiles in the Internet-based NutriNet-Sante Cohort. *Am J Epidemiol* 2014;179(7):910-916. [Medline: [24521560](https://pubmed.ncbi.nlm.nih.gov/24521560/)]
36. Vergnaud AC, Touvier M, Mejean C, Kesse-Guyot E, Pollet C, Malon A, et al. Agreement between web-based and paper versions of a socio-demographic questionnaire in the NutriNet-Sante study. *Int J Public Health* 2011;56(4):407-417. [Medline: [21538094](https://pubmed.ncbi.nlm.nih.gov/21538094/)]
37. Touvier M, Kesse-Guyot E, Mejean C, Pollet C, Malon A, Castetbon K, et al. Comparison between an interactive web-based self-administered 24 h dietary record and an interview by a dietitian for large-scale epidemiological studies. *Br J Nutr* 2011;105(7):1055-1064. [Medline: [21080983](https://pubmed.ncbi.nlm.nih.gov/21080983/)]
38. Freedman LS, Commins JM, Moler JE, Willett W, Tinker LF, Subar AF, et al. Pooled results from 5 validation studies of dietary self-report instruments using recovery biomarkers for potassium and sodium intake. *Am J Epidemiol* 2015;181(7):473-487. [Medline: [25787264](https://pubmed.ncbi.nlm.nih.gov/25787264/)]



39. Freedman LS, Commins JM, Moler JE, Arab L, Baer DJ, Kipnis V, et al. Pooled results from 5 validation studies of dietary self-report instruments using recovery biomarkers for energy and protein intake. *Am J Epidemiol* 2014;180(2):172-188. [Medline: [24918187](#)]
40. Crispim SP, Geelen A, Souverein OW, Hulshof PJ, Ruprich J, Dofkova M, EFCOVAL Consortium. Biomarker-based evaluation of two 24-h recalls for comparing usual fish, fruit and vegetable intakes across European centers in the EFCOVAL Study. *Eur J Clin Nutr* 2011 Jul;65 Suppl 1:S38-S47. [doi: [10.1038/ejcn.2011.86](#)] [Medline: [21731004](#)]
41. Øverby NC, Serra-Majem L, Andersen LF. Dietary assessment methods on n-3 fatty acid intake: a systematic review. *Br J Nutr* 2009 Dec;102 Suppl 1:S56-S63. [doi: [10.1017/S000711450999314X](#)] [Medline: [20100368](#)]
42. Dunn KM, Jordan K, Lacey RJ, Shapley M, Jinks C. Patterns of consent in epidemiologic research: evidence from over 25,000 responders. *Am J Epidemiol* 2004;159(11):1087-1094. [Medline: [15155293](#)]
43. Egan TM, Eide G, Gulsvik A, Bakke P. Nonresponse in a community cohort study: predictors and consequences for exposure-disease associations. *J Clin Epidemiol* 2002 Aug;55(8):775-781. [Medline: [12384191](#)]
44. Seybert H. Internet use in households and by individuals in 2012. Bruxelles european Union Eurostat 2012.
45. Huybrechts K, Mikkelsen E, Christensen T, Riis A, Hatch E, Wise L, et al. A successful implementation of e-epidemiology: the Danish pregnancy planning study 'Snart-Gravid'. *Eur J Epidemiol* 2010 May;25(5):297-304 [FREE Full text] [doi: [10.1007/s10654-010-9431-y](#)] [Medline: [20148289](#)]
46. Sautron V, Péneau S, Camilleri GM, Muller L, Ruffieux B, Hercberg S, et al. Validity of a questionnaire measuring motives for choosing foods including sustainable concerns. *Appetite* 2015 Apr;87:90-97. [doi: [10.1016/j.appet.2014.12.205](#)] [Medline: [25529817](#)]
47. Camilleri GM, Méjean C, Bellisle F, Andreeva VA, Sautron V, Hercberg S, et al. Cross-cultural validity of the Intuitive Eating Scale-2. Psychometric evaluation in a sample of the general French population. *Appetite* 2015 Jan;84:34-42. [doi: [10.1016/j.appet.2014.09.009](#)] [Medline: [25239401](#)]

## Abbreviations

**BMI:** body mass index

**DHA:** docosahexaenoic acid

**EPA:** eicosapentaenoic acid

**ICCs:** intraclass correlation coefficients

*Edited by G Eysenbach; submitted 22.04.16; peer-reviewed by D Thompson, J Torous; comments to author 25.07.16; revised version received 25.08.16; accepted 28.09.16; published 18.10.16.*

*Please cite as:*

*Kesse-Guyot E, Assmann K, Andreeva V, Castetbon K, Méjean C, Touvier M, Salanave B, Deschamps V, Péneau S, Fezeu L, Julia C, Allès B, Galan P, Hercberg S*

*Lessons Learned From Methodological Validation Research in E-Epidemiology*

*JMIR Public Health Surveill* 2016;2(2):e160

URL: <http://publichealth.jmir.org/2016/2/e160/>

doi: [10.2196/publichealth.5880](#)

PMID: [27756715](#)

©Emmanuelle Kesse-Guyot, Karen Assmann, Valentina Andreeva, Katia Castetbon, Caroline Méjean, Mathilde Touvier, Benoît Salanave, Valérie Deschamps, Sandrine Péneau, Léopold Fezeu, Chantal Julia, Benjamin Allès, Pilar Galan, Serge Hercberg. Originally published in *JMIR Public Health and Surveillance* (<http://publichealth.jmir.org>), 18.10.2016. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in *JMIR Public Health and Surveillance*, is properly cited. The complete bibliographic information, a link to the original publication on <http://publichealth.jmir.org>, as well as this copyright and license information must be included.



Original Paper

# Exploring Consumer and Patient Knowledge, Behavior, and Attitude Toward Medicinal and Lifestyle Products Purchased From the Internet: A Web-Based Survey

Sulaf Assi<sup>1\*</sup>, PhD; Jordan Thomas<sup>1</sup>, BSc.; Mohamed Haffar<sup>2\*</sup>, PhD; David Osselton<sup>1\*</sup>, PhD

<sup>1</sup>Faculty of Science and Technology, Department of Archaeology, Anthropology and Forensic Science, Bournemouth University, Poole, United Kingdom

<sup>2</sup>Faculty of Management, Bournemouth University, Poole, United Kingdom

\*these authors contributed equally

**Corresponding Author:**

Sulaf Assi, PhD

Faculty of Science and Technology

Department of Archaeology, Anthropology and Forensic Science

Bournemouth University

Christchurch House, Room C213

Fern Barrow

Poole, BH12 5BB

United Kingdom

Phone: 44 1202961264

Fax: 44 1202 961179

Email: [sassi@bournemouth.ac.uk](mailto:sassi@bournemouth.ac.uk)

## Abstract

**Background:** In recent years, lifestyle products have emerged to help improve people's physical and mental performance. The Internet plays a major role in the spread of these products. However, the literature has reported issues regarding the authenticity of medicines purchased from the Internet and the impact of counterfeit medicines on public health. Little or no data are available on the authenticity of lifestyle products and actual toxicity associated with their use and misuse.

**Objective:** Our aim was to investigate consumer and patient attitudes toward the purchase of lifestyle products from the Internet, their knowledge of product authenticity and toxicity, and their experiences with counterfeit lifestyle products.

**Methods:** A Web-based study was performed between May 2014 and May 2015. Uniform collection of data was performed through an anonymous online questionnaire. Participants were invited worldwide via email, social media, or personal communication to complete the online questionnaire. A total of 320 participants completed the questionnaire.

**Results:** The results of the questionnaire showed that 208 (65.0%) participants purchased lifestyle products from the Internet mainly due to convenience and reduced cost. More than half (55.6%, 178/320) of participants purchased cosmetic products, whereas only a minority purchased medicinal products. Yet, 62.8% (201/320) of participants were aware of the presence of counterfeit lifestyle products from the Internet, and 11.9% (38/320) experienced counterfeit products. In only 0.9% (3/320) of those cases were counterfeit lifestyle products reported to authorities. Moreover, 7.2% (23/320) of the participants experienced adverse effects due to counterfeit lifestyle products.

**Conclusions:** In summary, patients experienced counterfeit lifestyle products that resulted in adverse effects on their health. Although certain adverse effects were reported in this study, counterfeit products were underreported to authorities. Further public awareness campaigns and patient education are needed.

(*JMIR Public Health Surveill* 2016;2(2):e34) doi:[10.2196/publichealth.5390](https://doi.org/10.2196/publichealth.5390)

**KEYWORDS**

consumer knowledge; attitude; behaviour authenticity; effectiveness; toxicity; counterfeit; medicinal and lifestyle products; Internet

## Introduction

The last few years have witnessed a major change in perspectives toward medicine. Whereas medicine had been previously utilized for lifesaving purposes, a new era has emerged that involves the use of medicine to improve the overall lifestyle of individuals. In this respect, lifestyle products are intended mainly for improvement in mental functions and physical performance rather than curing diseases [1,2]. Lifestyle products can be of medicinal or non-medicinal origin, of any formulation type (eg, tablets, powders, creams, solutions), and any source (eg, herbal, synthetic). These products are classified into variable categories that improve mood and social behavior, cognitive functions, physical appearance, and/or sexual performance [1].

The Internet plays a major role in the dissemination of these products with many advantages over traditional marketplaces [3-6]. In this sense, the Internet offers a quick, easy, and more convenient way for purchasing medicinal and health care products. Specifically for medicinal products, the Internet offers regulated medicines without the need for a prescription [5,7]. Furthermore, Internet orders can be placed from home and at any hour of the day. Also, privacy is preserved with online purchases compared to face-to-face purchases [8], saving the consumer potential embarrassment. Additionally, the Internet provides more detailed information about the products and reduces visits to health care professionals and community pharmacies [7].

However, the issues associated with lifestyle products (including medicinal, herbal, dietary supplements, and cosmetics) purchased from the Internet are much more complicated. The purity and quality of these products represents a major concern and impact on consumer health [9-11]. For instance, these products may be defective in their packaging and ingredients. They may be poorly stored or past their best-before date. For instance, amphetamines and ephedrine encountered in counterfeit herbal weight loss products could result in sympathomimetic side effects such as hyperthermia, hypertension, and agitation [12,13]. Furthermore, heavy metals and pesticide contamination in counterfeit products often results in both acute and chronic side effects [14].

According to the World Health Organization, more than 50% of medicines purchased online could be counterfeit [15]. A further study conducted in the United States in 2004 showed that more than 80% of the medicines purchased from the Internet contained the wrong active pharmaceutical ingredient (API), were subpotent, out-of-date, or poorly stored [16]. Drug websites, unlicensed online pharmacies, and other unregulated online retailers have been found to sell counterfeit/substandard products [17,18]. In the United States, the Food and Drug Administration (FDA) warns against "rogue websites" that sell potentially dangerous drugs [19]. According to the FDA, these drugs may contain the wrong API, too much or too little API, or even dangerous API. Moreover, the National Association of Boards of Pharmacy specifies that more than 10,000 Internet websites selling medicines do not adhere to the pharmacy and practice standards [20].

Counterfeit products encountered on the Internet comprise a diverse range of pharmacological classes and formulations. Thus, Interpol carries out yearly operations in order to tackle the sales of counterfeit medicines online [21]. In 2015, operation Pangea VIII involved the seizures of 20.7 million counterfeit and illicit medicines including antihypertensive medicines, anticancer agents, sexual stimulants, and nutrition supplements [21]. Additional counterfeit products found on the Internet in other studies included anabolic steroids, anticancer, antiviral, antidepressants, anxiolytics, contraceptive, sexual stimulants, and weight-loss medicines [19,20,22-24]. In this case, anabolic steroids, along with anxiolytics and weight-loss pills sold on the Internet, were seized in the United Kingdom [22]. Anticancer agents encountered in the United States were Avastin in 2012 [20]. Additionally, Tamiflu purchased online was shown to contain a mixture of paracetamol and talc instead of its API (oseltamivir) [19]. Also, a number of antidepressants/anxiolytics (ie, alprazolam, escitalopram, lorazepam, and zolpidem) purchased online were found to not contain their API and instead contain another antipsychotic (haloperidol) [19]. Moreover, contraceptive products purchased from the Internet under the brand "Ortho Eva" did not contain any API [23]. Counterfeit sexual stimulants sold on the Internet, mainly Viagra, have been reported in a number of studies [24]. Also, sibutramine was encountered instead of orlistat in counterfeit Xenical product [19].

The harm resulting from using counterfeit medicinal products could range from ineffectiveness to potentially lethal effects. It is noteworthy to mention that the harm resulting from counterfeit lifestyle products could be a great risk in both medicinal and non-medicinal products. For instance, toxic metals were detected in both counterfeit cosmetic and herbal products [25-27]. Nonetheless, the majority of the literature investigated harm associated with counterfeit medicinal products. Such harm was based on predicting the adverse effects of certain medicines or experience of severe toxicity/lethal effect with a counterfeit medicine. Only four studies in the literature evaluated pharmacist [28] and patient/consumer knowledge [4,6,29] of medicines, and/or herbal products and dietary supplements. However, these studies concentrated mainly on products used for medicinal purposes and did not evaluate additional lifestyle products (eg, cosmetics) used by consumers on a daily basis. Moreover, they did specify the degree of harm (ie, mild, moderate, or severe) resulting from the use of counterfeit lifestyle products.

Our work aimed to investigate consumer and patient attitudes toward the purchase of lifestyle products from the Internet, their knowledge about product effects, authenticity, and toxicity, as well as their experience with using counterfeit lifestyle products.

## Methods

### Study Design and Data Collection

A multinational Web-based study was designed in order to examine the knowledge and attitudes of patients and consumers toward lifestyle products sold on the Internet. Participants were eligible if they spoke English and purchased lifestyle products from the Internet. Participants who did not make any Internet

purchases were excluded. Residents in 22 countries responded, including Belgium, Brazil, Canada, China, Finland, France, Germany, Hong Kong, Italy, Kuwait, Lebanon, Morocco, Pakistan, Palestine, Russia, Saudi Arabia, Thailand, Turkey, United Arab Emirates, United Kingdom, and United States.

Data were collected using an anonymous online questionnaire by Bristol Online Survey and were accessible for the lifetime of the questionnaire (ie, between May 2014 and May 2015). Participants worldwide were invited to complete the survey by sending the link through emails and personal communication. Also, the questionnaire was posted on social media websites and discussion forums (eg, Facebook and LinkedIn). The language of the questionnaire was English (see [Multimedia Appendix 1](#)).

The questionnaire was initially piloted at a local gym (n=15) and at the university (n=15) prior to the study in order to

evaluate reliability and clarity of the information using Cronbach alpha. After 1 week, the questionnaire was retested for reliability with the same 30 participants. A few modifications were made based on the outcomes of the pilot study, and the final version of the questionnaire was used online to collect data.

## Questionnaire

The questionnaire was divided into six parts that covered the following areas (see [Table 1](#)): (1) demographics, (2) extent of buying lifestyle products from the Internet, (3) types of products purchased, (4) awareness of counterfeit products sold over the Internet, (5) experience with counterfeit lifestyle products, and (6) experience of harm (adverse effects) associated with the use of counterfeit or poor-quality lifestyle products ([Multimedia Appendix 1](#)).

**Table 1.** Areas covered by the questionnaire.

Part	Aims
Demography	Sociodemographic Economic
Extent of buying products from the Internet	Frequency of buying products Sources of products Details of online retailers/pharmacies
Types of the products purchased	Medicinal products Non-medicinal products (herbal, supplements, and cosmetics) Classes of products (image improvement or performance enhancement)
Awareness of counterfeit products sold over the Internet	Knowledge on counterfeit products Sources of knowledge
Experience with counterfeit lifestyle products	Experience with adverse effects Types of adverse effects Degree of harm

## Ethics

Ethical approval for the study was sought by Bournemouth University internal ethics committee. Moreover, respondents gave informed consent of their willingness to take part in this study at the introduction of the questionnaire. The introduction further clarified to the respondents that they could withdraw anytime from the study ([Multimedia Appendix 1](#)). Moreover, all respondents' data were handled and stored anonymously.

## Definitions

Lifestyle products are those intended to enhance the physical appearance and/or physical/mental performance of individuals [1]. A counterfeit mark is defined as “a spurious mark which is identical with or is substantially indistinguishable from a registered mark” [30]. A counterfeit medicinal product is defined as medicine that is “fraudulently and deliberately mislabeled according to identity and/or source” [31]. A counterfeit medicinal product could contain no API, wrong API, wrong ingredients, or even defective packaging [31]. A cosmetic product is [32]:

*Any substance or preparation intended to be placed in contact with various external parts of the human body (epidermis, hair system, nails, lips and external genital organs) or with the teeth and the mucous membranes of the oral cavity with a view exclusively or mainly to cleaning them, perfuming them, changing their appearance and/or correcting body odors and/or protecting them or keeping them in good condition.*

## Data Analysis

Data analysis was conducted using SPSS v21 where descriptive statistics were applied in order to gather responses and explore outcomes. Moreover, responses from open-ended questions were investigated individually in relation to consumer and patient knowledge and attitudes toward lifestyle products, their authenticity, and the associated toxicity with counterfeit lifestyle products. As most data obtained were categorical variables, they were reported as numbers and frequencies.

## Results

The questionnaire yielded 320 respondents from different sources. Participants were invited to complete the questionnaire online. The responses were received via the website, and all were complete and usable.

### Sociodemographics

Information from 320 respondents was analyzed. The respondents included 91 females (28.4%), 227 males (70.9%), and 2 participants who did not disclose their gender (see [Table 2](#)). The majority (62.5%, 200/320) of the respondents were in the 18-25 years group, 47 (14.7%) in the 26-33 years group, and 43 (13.4%) in the 34-41 years group. The remaining age groups were represented in less than 10% of the patients. Most

of the respondents were British (78.1%, 250/320), followed by Europeans (8.8%, 28/320), Asian (7.5%, 24/320), African (1.3%, 4/320), Australians (0.9%, 3/320), and Americans (0.63%, 2/320). In addition, the majority of the respondents were residents of the United Kingdom (85.0%, 272/320), Asia (6.3%, 20/320), and Europe (3.4%, 11/320). The educational level among the respondents was mainly at a higher degree level or above; 149 (46.6%) of the respondents had at least a Bachelor's degree. In relation to the number of languages spoken among respondents, 88.4% (283/320) were monolingual and spoke English only. The remaining respondents were bilingual or trilingual and spoke the following languages in addition to English: Arabic, Danish, Dutch, French, German, Greek, Italian, Mandarin, Nepalese, Russian, Somali, Spanish, Swedish, Thai, Turkish, and Urdu.

**Table 2.** Sociodemographic characteristics of participants (N=320).

Parameter	n	Frequency, %
<b>Age group</b>		
18-25	200	62.5
26-33	47	14.7
34-41	29	9.06
42+	43	13.4
Prefer not to say	1	0.31
<b>Gender</b>		
Male	91	28.4
Female	227	70.9
Prefer not to say	2	0.63
<b>Nationality</b>		
Africa (Ivorian, Moroccan, Nigerian, South African)	4	1.25
Asia (Bangladeshi, Egyptian, Indian, Kuwaiti, Lebanese, Nepal, Palestinian, Pakistani, Russian, Saudi Arabian, Syrian, Thai, and Turkish)	24	7.5
Australian	3	0.94
Europe (British)	250	78.1
Europe other (Belgian, Bulgarian, Columbian, Cypriot, Czech, Danish, Dutch, Finnish, French, German, Greek, Italian, Norwegian, Polish, Romanian, Spanish, Swedish)	28	8.75
North America (American and Canada)	8	2.50
South America (Brazilian, Colombian)	2	0.63
<b>Country of residence</b>		
Africa	0	0
Asia (China, Hong Kong, Kuwait, Lebanon, Morocco, Thailand, Turkey, Pakistan, Palestine, Russia, Saudi Arabia, United Arab Emirates)	20	6.25
Australia	3	0.94
Europe (UK)	272	85
Europe (Belgium, Finland, France, Germany, Italy)	11	3.4
North America (America and Canada)	13	4.06
South America (Brazil)	1	0.31
<b>Second language (additional to English)</b>		
None	283	88.4
One second language (Arabic, Danish, Dutch, French, German, Greek, Italian, Nepalese, Russian, Somali, Spanish, Swedish, Thai, Turkish, Urdu)	17	5.31
<b>Education level</b>		
School/College	109	34.1
Bachelor's degree	149	46.6
Master or post graduate	44	13.8
PhD+	18	5.63



### Consumer and Patient Attitude Toward the Purchase of Lifestyle Products

According to the survey, 208 (65.0%) of the respondents reported purchasing lifestyle products from the Internet (see [Table 3](#)). However, only 17 (5.3%) claimed to have purchased lifestyle products frequently, whereas 111 (34.7%) purchased products occasionally and 91 (28.4%) rarely purchased these products. The main websites used for purchase of products were Amazon (64.1%, 205/320), eBay (39.7%, 127/320), online retailers' websites (32.8%, 105/320), and online pharmacies (14.7%, 47/320). On the other hand, a very small proportion (<10%) used a "drug" website or Alibaba. Retailers websites reported included (1) cosmetic retailers: All Beauty, Alvin Connor, Bodyshop, Dermashop, MAC make up, (2) health care and beauty online retailers: Boots, Healthspan, Holland and Barrett, Love Melatonin, My Protein, Superdrug, (3) online department stores: Debenhams, Feel Unique, John Lewis, QVC, (4) supermarkets, (5) other websites such as Groupon, and (6) wholesale South Asian suppliers. When asked whether the country of the website was identifiable, fewer than half (44.7%, 143/320) of respondents could identify the countries of origin of the websites, which were France, Germany, India, Japan,

United Kingdom, and United States. Some respondents claimed they could identify countries of origin of online pharmacies, which were mainly the United Kingdom (23.8%, 76/320) and United States (5.3%, 17/320). Other countries reported as sources for online pharmacies were Australia, Canada, China, Czech Republic, Finland, Germany, Greece, Holland, India, Pakistan, Philippine, Singapore, Sweden, Switzerland, and Thailand. Among the respondents who already bought medicines from UK online pharmacies, only 53 (16.6%) could recognize the Medicine and Healthcare Regulatory Agency (MHRA) logo. Moreover, only 17 (5.3%, 17/320) of respondents had consultations with a doctor at the online pharmacy. Subsequently, the sources of information obtained by the respondents regarding their products were mainly from the Internet (45.3%, 145/320) or family/friends (43.4%, 139/320) (see [Table 3](#)). Internet sources reported to be used by respondents were product websites, blogs/forums/chat rooms, YouTube, scientific papers, Facebook, and NHS (National Health Service) Direct website. In addition, a lower percentage of respondents obtained information on lifestyle products from interacting with health care professionals (24.1%, 77/320), magazines (17.8%, 57/320), and TV (13.1%, 42/320).

**Table 3.** Consumer and patient attitudes toward purchasing lifestyle products from the Internet.

Criteria	n	Frequency, %
<b>Purchase of lifestyle products over the Internet</b>		
Yes	208	65
No	112	35
<b>Frequency of purchase</b>		
Always	17	5.31
Occasionally	111	34.7
Rarely	91	28.4
Never	101	31.6
<b>Websites mostly used</b>		
Alibaba	7	2.19
Amazon	205	64.1
eBay	127	39.7
Drugs websites	21	6.56
Online pharmacies	47	14.7
Others (All Beauty, Alvin connor, Beauty base, Bodyshop, Boots, Debenham, Dermashop, Feel unique, Groupon, Healthspan, Holland and Barrett, John Lewis, Love Melanotan, MAC makeup, My protein,QVC, Superdrug, Tesco, Wholesale Indian suppliers)	105	32.8
<b>Country of the website identifiable</b>		
Yes (France, Germany, India, Japan, UK, and US)	143	44.7
No	103	32.2
N/A	74	23.1
<b>Country of origin for medicines bought from online pharmacies</b>		
UK	76	23.8
USA	17	5.31
India	6	1.88
Canada	4	1.25
Germany	4	1.25
Australia	2	0.63
Holland	2	0.63
Other countries one pharmacy in each (China, Czech Republic, Finland, Greece, Pakistan, Philippines, Singapore, Sweden, Switzerland, Thailand)	10	3.13
<b>MHRA logo for authenticity for UK online pharmacies</b>		
Yes	53	16.6
No	30	9.38
N/A	237	74.1
<b>Consultation with a doctor at the online pharmacy</b>		
Yes	17	5.31
No	115	35.9
N/A	188	58.8
<b>Sources of information about lifestyle products</b>		
Family/friends	139	43.4
Health care professionals	77	24.1
Magazines	57	17.8

Criteria	n	Frequency, %
TV	42	13.1
Others (websites, forums/blogs, Internet ads, smartphone apps, YouTube, Scientific papers, Product information leaflet, Facebook, NHS Direct)	145	45.3

### Types of Lifestyle Products Purchased

The results of the questionnaire showed that the main lifestyle products purchased from the Internet were cosmetic products (55.6%, 178/320) representing more than half of the total products (see Table 4). This was followed by supplementary/nutritional lifestyle products, drugs and herbal products, which represented 28.8% (92/320) and 18.8% (60/320) of the total purchased products respectively. Medicinal products represented the lowest percentage of the purchased lifestyle products (11.9%, 38/320).

Half the respondents 171 (53.4%) claimed that the lifestyle products they purchased from the Internet were available in stores in main shopping areas. They gave several reasons that made them purchase products online. The main justifications were convenience (25.9%, 83/320), lower cost/better offers (25.9%, 83/320), and easier alternative to in-store purchase (15%, 48/320). However, very few (4.7%, 15/320) participants believed Internet purchases were timesaving and quick, and only 6 (1.9%, 6/320) said that Internet purchases saved them from the embarrassment that could be encountered in stores when buying personal products. Also, less than 1% of respondents reported that Internet purchases provided a wider variety of products and offered more details on the products than those given by the pharmacist/store assistant. Even fewer reported that they favored Internet purchases because of better-quality products and ability to buy prescription medicines without a prescription.

Cosmetic products used for skin and physical appearance were the most commonly purchased products and were used by more than 60% of respondents. This was followed by lifestyle products used for mood and social behavior and those used for cognitive function, which were used by 34.7% (111/320) and 29.1% (93/320) respondents. Weight loss products and sexual stimulants were used by 20-30% of the respondents, while muscle enhancers were used by 16.3% (52/320) of respondents.

More specifically, the subclasses reported in each product category varied between each category. For cosmetic products, the majority of products used were hair products (n=220), acne products (n=149), moisturizers (n=140), sunscreens (n=55), and tanning solutions (n=43). In the remaining categories, herbal products formed the main category purchased. Products purchased in relation to mood and social behavior included mainly herbal sleep aids (n=50), mood enhancers (n=19), stimulants (n=19), and antidepressants (n=18). Cognitive enhancers were mainly caffeine (n=62) and natural memory enhancers (n=20). Only 7 respondents claimed the use of nootropics (synthetic medicines) for cognitive function. Similarly, synthetic medicines were underrepresented in the sexual stimulant category. Thus, the main purchases in this category were condom-type products, whereas only 14 (4.4%) respondents reported the purchase of Viagra, Cialis, and Levitra tablets. Herbal weight loss products (n=42) were preferred over other synthetic appetite suppressants (n=20), fat binders (n=16), or meal replacements (n=6). Furthermore, proteins (n=49) were favored over steroids (n=3) for muscle enhancers.

**Table 4.** Types of lifestyle products purchased from the Internet.

Criteria	n	Frequency, %
<b>Type of products purchased online</b>		
Cosmetic	178	55.6
Herbal	60	18.8
Medicinal	38	11.9
Supplementary and nutrition	92	28.8
Others (drugs/legal highs, essential oils, condoms)	88	27.5
<b>Availability of products ordered online in community pharmacies/stores</b>		
Yes	171	53.4
No	149	46.6
<b>Reasons for buying products online</b>		
Quick/time saving	15	4.69
Cheap	83	25.9
Easy	48	15
Convenient	83	25.9
More details on product than ones provided by the pharmacist	2	0.63
No need for embarrassment of communicating with pharmacist	6	1.88
Lack of availability of products in stores	9	2.81
Not able to get prescription for some products	3	0.94
Wide variety of products and offers	2	0.63
Ability to purchase illegal drugs	1	0.31
Better quality products	1	0.31
<b>Cognitive function</b>		
Caffeine	62	19.4
Natural memory enhancers	20	6.25
Nootropics	7	2.19
Others (multivitamins, cod liver oil, stimulant drugs)	4	1.25
Total	93	29.1
<b>Mood and social behavior</b>		
Antidepressants	18	5.63
Mood enhancers	19	5.94
Sleep aids	50	15.63
Highs	19	5.94
Others (psychedelics, anxiolytics, addiction management, pain management)	5	1.56
Total	111	34.7
<b>Physical appearance</b>		
Acne products		
Creams/ointments/gels	95	29.7
Solutions	12	3.75
Tablets	42	13.1
Hair products		
Hair loss	9	2.81
Cream/gel	48	15

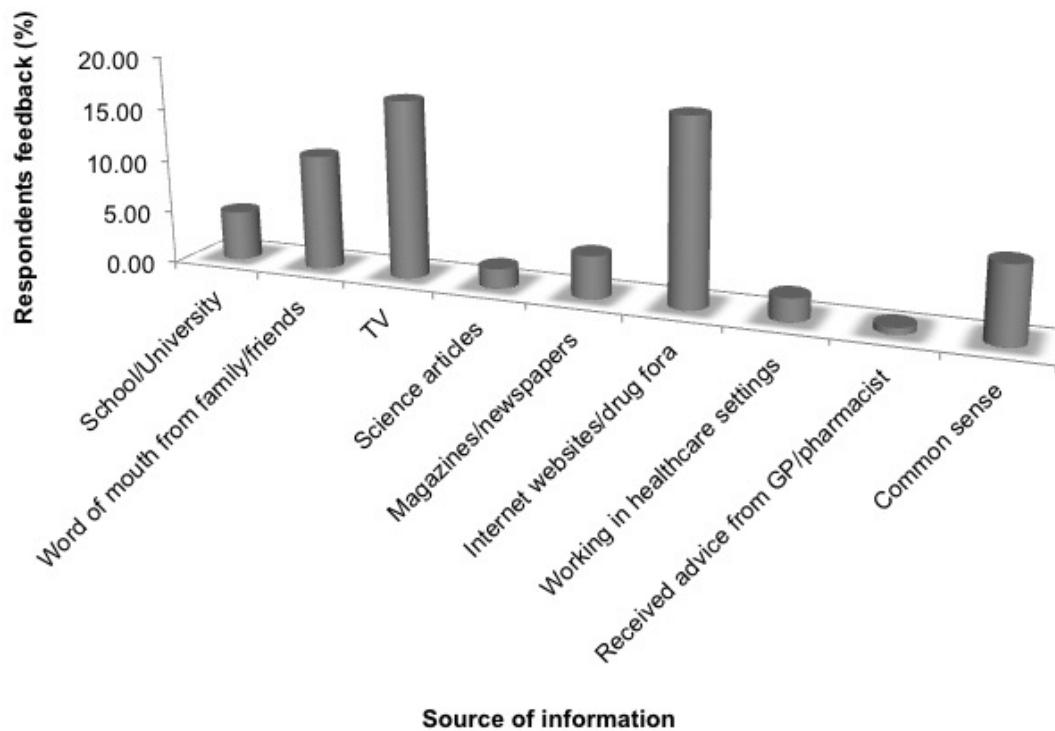
Criteria	n	Frequency, %
Shampoo	96	30
Tablet	17	5.31
Hair dye	50	15.6
Other (nail polish, mascara, deodorant, perfume, lip balm, herbal products, protein shakes, make-up, shampoo)	17	5.31
Total	237	74.1
<b>Muscle enhancers</b>		
Proteins	49	15.3
Steroids	3	0.94
Total	52	16.3
<b>Skin products</b>		
Moisturizers	140	43.8
Skin lighteners	10	3.13
Sunscreens	55	17.2
Tanning solutions	43	13.4
Other (foundation, body scrub, body oil)	5	1.56
Total	253	79.1
<b>Weight-loss products</b>		
Appetite suppressants	20	6.25
Fat binders	16	5
Herbal products	42	13.1
Other (juices and meal replacement, diet shakes, diet pill)	6	1.88
Total	84	26.3
<b>Sexual stimulants</b>		
Condom products	44	13.8
Herbal products	12	3.75
Libido enhancers	6	1.88
Viagra	7	2.19
Cialis	3	0.94
Levitra	4	1.25
Total	76	23.8

### Awareness of Counterfeit Lifestyle Products Sold Via the Internet

Most respondents (62.8%, 201/320) were aware of counterfeit lifestyle products being sold via the Internet. However, the sources of information regarding product counterfeiting varied between respondents (see [Figure 1](#)). In the majority of cases, respondents relied on information from Internet websites/drug forums (17.5%, 56/320), TV (16.9%, 54/320), and family/friends (10.9%, 35/320). Also, 25 (7.2%) respondents claimed that awareness of counterfeits is “common sense.” Moreover, 13

(4.1%) respondents gained knowledge about counterfeiting from magazines and newspapers. Furthermore, 7 (2.2%) respondents knew about counterfeiting from their job within a health care setting. Education was not a major source, as 15 (4.7%) respondents claimed they learned about counterfeiting in school/university degree and only 6 (2.0%) read scientific articles on the topic. Additionally, advice received from health care professionals was not enough regarding counterfeiting as only 2 respondents reported that their general practitioner and pharmacist explained product counterfeiting to them.



**Figure 1.** Sources of information obtained by consumers regarding counterfeit lifestyle products.

1

### Experience of Counterfeit Lifestyle Products

When asked about experience with a counterfeit product, 38 (11.9%) respondents reported having experience with a counterfeit lifestyle product (see [Table 5](#)). In only 3 cases was this product reported to an authority. Moreover, when asked how they could identify a counterfeit product, respondents claimed that they would look at the product's packaging (12.8%, 41/320), appearance (11.6%, 37/320), or label claim (5.6%, 18/320). Fewer respondents stated that they would know from

the product's efficacy (2.5%, 8/320) or side effects (4.1%, 13/320).

Respondents were aware of the risks associated with the counterfeit products. However, 45 (14.1%) respondents said that it was acceptable to take the risk of buying potential counterfeit products in case of emergencies such as medicine shortage or poor finance. They further clarified that it is up to the individual who is aware of the problem to make the decision especially in case of poor finance, lack of availability in online stores, or "if the product is not a medicine."

**Table 5.** Consumer experience with counterfeit lifestyle products.

Criteria	n	Frequency, %
<b>Experience of a counterfeit products</b>		
Yes	38	11.9
No	282	88.1
<b>Identification of a counterfeit product</b>		
Different appearance	41	12.8
Different packaging	37	11.6
No label	18	5.63
No packaging	10	3.13
Side effects	13	4.06
Wrong ingredients	8	2.50
Other (not effective, different brand printed on the label, adverse effects not stated on the product information leaflet, cosmetics giving lighter colors)	7	2.19
<b>Was the counterfeit product reported to the authority</b>		
Yes	3	0.94
No	99	30.9
Other (reported to eBay)	1	0.31
<b>Risking buying a potential counterfeit product from the Internet due to medicine shortage or lower prices</b>		
Yes	45	14.1
No	249	77.8
Not applicable	26	8.13
<b>Reasons for risking buying a potential counterfeit product from the Internet due to medicine shortage or lower prices</b>		
Cheaper	7	2.19
Easier	3	0.94
Lack of availability in stores	4	1.25
Flexibility	1	0.31
Up to individual after becoming aware of the problem	4	1.25
Depends on the product (okay for cosmetics but not for drugs)	4	1.25

### Experience of Adverse Effects Resulting From Counterfeit Lifestyle Products

Adverse effects resulting from the use of counterfeit products were below 10% (see Table 6); only 23 (7.2%) respondents claimed they had adverse effects resulting from a counterfeit product purchased from the Internet. Of these, 17 reported that the adverse effects were not stated on the product information leaflet and were instead due to the poor quality. The types of adverse effects varied between products and were mainly encountered with cosmetic and herbal products. Skin reactions

(eg, allergy, rash, itching, swelling) as well as eye infection were experienced with counterfeit cosmetic products. In addition, increased blood pressure/heart rate, appetite suppression, urinary tract infection, and gastrointestinal disturbances were reported with herbal products. Only 4 respondents reported these adverse effects to the authorities/source of purchase, and only 1 respondent received treatment for these adverse effects. However, about half of respondents (49.4%, 158/320) believed that the extent of harm resulting from counterfeit products could be lethal.

**Table 6.** Consumers' experience with adverse reactions associated with the use of counterfeit lifestyle products.

Criteria	n	Frequency, %
<b>Experience of an adverse effect from products purchased online</b>		
Yes	23	7.19
No	297	92.81
<b>Were the side effects stated in the products' information leaflet</b>		
Yes	17	5.31
No	258	80.6
No label	45	14.1
<b>Types of adverse effects experienced</b>		
Skin reactions (allergy, rash, itching, swelling)	15	4.69
Increased blood pressure	2	0.63
Increased heart rate	2	0.63
Eye infection	1	0.31
Urinary tract infection	1	0.31
Nausea, vomiting, gastrointestinal disturbances	8	2.5
Appetite suppression	1	0.31
<b>Reported adverse effects</b>		
Yes	4	1.25
No	143	44.7
<b>Receipt of treatment for adverse effects</b>		
Yes	1	0.31
No	22	6.88
<b>Extent of harm resulting from counterfeit product</b>		
Very mild	15	4.69
Mild	33	10.3
Average	31	9.69
Harmful	83	25.9
Lethal	158	49.4

## Discussion

### Principal Findings

This study examined consumer and patient attitudes toward the purchase of lifestyle products from the Internet. Although previous literature surveys examined the attitudes of patients towards online pharmacies [4,6,29], they did not evaluate products such as cosmetics and herbal medicines that could impact public health or patient safety. To our knowledge, this was the first study to address user perceptions regarding counterfeit lifestyle products. Lifestyle products have witnessed a global increase in recent years due to the change in people's attitude, way of living [2], and the introduction of personalized care [1,11]. Personalized care consists of licensed/unlicensed products that focus mainly on improving an individual's

performance, image, mood, appetite, sleep, and sexual desires [1]. Moreover, these products were found to be sold on nearly every website in a study evaluating 136 online pharmacies [11].

The research showed that more than half of respondents purchased lifestyle products from the Internet. This was much higher than previous investigations that reported 8.3% [6], 14.5% [29], and 16% [4] of consumers bought medicinal products from the Internet. This could be attributed to the fact that the latter investigations evaluated only medicines, whereas our study included medicines, cosmetics, and herbal products. This would imply that additional lifestyle products were underestimated in previous studies despite their impact on public health. Moreover, the aforementioned surveys were limited to participants from a small range of countries, whereas this survey had respondents of 40 nationalities from 22 countries.

The majority of participants spoke only English, and a minority of them spoke two or more languages. The majority of the participants were UK residents, which influenced the types of lifestyle products purchased. Amazon and eBay were the most popular sites mentioned by respondents. This was followed by known health care and beauty retailer websites such as Boots and Superdrug. In contrary to other studies [4], online pharmacies were used by only 14% of participants. Less than half of participants were aware of the country of origin of the website/online pharmacy they were buying from. Specifically for online pharmacies, only 16% could recognize the MHRA logo for licensed online pharmacies. Also, respondents did not communicate with cyber doctors for information on lifestyle products. However, they did count mainly on family and friends, the media, or the Internet. Only a quarter of respondents obtained information from health care professionals. This was also noted in other studies that showed that the lack of communication between health care professionals and patients affected the desired treatment outcome [33].

It is likely that respondents did not consult with health care professionals because more than half of the products purchased from the Internet were cosmetics. Though the products were available in stores, respondents preferred purchasing them from the Internet as it offered quicker, cheaper, and more convenient alternatives to stores [4,33,34]. The cosmetics were mainly hair/face products, acne products, and tanning solutions. Additionally, respondents used products that improved their mental performance (eg, mood, social behavior, and cognitive function) and physical appearance (eg, weight loss products, sexual stimulants, and muscle enhancers).

The survey found that more than half of respondents were aware of counterfeit products. Respondents referred to Internet websites, media, and family/friends as the major source of information on counterfeit products; however, half of them did not check the origin of the website when purchasing lifestyle products. This observation was in agreement with other surveys that showed that consumers relied on information obtained from the Internet regarding their products [35,36].

In only 1% of cases were counterfeit products reported to the authorities or the supplier. Some respondents did not report counterfeit products and underestimated their dangerous consequences. They believed it acceptable to risk buying counterfeit products in the case of medicine shortage, poor finance, lack of product availability in stores, or in cases of non-medicinal products. This was because of the perception that non-medicinal products (such as cosmetics) were considered less dangerous and harmful to the consumer.

The adverse effects experienced from the use of lifestyle products in this survey were relatively low (<10%). In most

cases, adverse effects were attributed to the product itself and corresponded to effects described in the patient information leaflet (PIL). On the contrary, adverse effects not stated on the PIL were attributed to the product being potentially counterfeit and included skin reactions (eg, allergy, rash, itching, and swelling), cardiovascular effects (eg, increased heart rate and blood pressure), gastrointestinal effects (eg, nausea, vomiting, and gastrointestinal disturbance), eye infection, and urinary tract infection. Yet in only 1% of the cases were these adverse effects reported to authorities or a treatment sought. However, more than three-quarters of participants believed that counterfeit medicines could be harmful or lethal.

### Limitations

The first limitation in this study was our sample size of 320 participants. The low response rate could be attributed to the complexity of the survey and can be improved with further development and testing. More specifically, the study sample size was low in some countries including Australia, China, Czech Republic, Finland, Greece, Holland, Pakistan, Philippine, Singapore, Sweden, Switzerland, and Thailand. The majority of responses were from the United Kingdom. In this respect, caution should be taken in interpreting the generalizability of the findings in the aforementioned countries. Moreover, the questionnaire was self-reported; thus, there could be potential information bias. Furthermore, the research is not itself conclusive and more research is needed to explore the association between counterfeit products and side effects experienced.

### Conclusions

The results of this study showed that more than half the respondents purchased lifestyle products from the Internet. The majority of the respondents purchased cosmetics; whereas, only a small minority purchased medicines from the Internet. The main reasons attributed to purchasing lifestyle products from the Internet were convenience, low prices, detailed product information, and consumer privacy. Most respondents were aware of the presence of counterfeit lifestyle products on the Internet. The main source of information about counterfeit products was obtained through media and/or family and friends. However, only 11.9% of patients experienced counterfeit lifestyle products, of whom only 1% reported it to authorities. Only 7.2% of patients experienced adverse effects associated with the use of a counterfeit lifestyle product. Nonetheless, all respondents were aware of the dangers associated with the use of counterfeit lifestyle products. However, 14.1% of respondents considered it acceptable to risk buying counterfeit products in the case of poor finance, lack of availability in store, or for non-medicinal products.

### Authors' Contributions

SA participated in the study design, questionnaire development and distribution, data analysis, and manuscript preparation. JT participated in the questionnaire development, questionnaire distribution, and data collection. MH participated in the questionnaire development and distribution. DO participated in the questionnaire development and quality checks of the manuscript. All authors have read and approved the final manuscript.

## Conflicts of Interest

None declared.

## Multimedia Appendix 1

Online survey questionnaire.

[[PDF File \(Adobe PDF File\), 62KB - publichealth\\_v2i2e34\\_app1.pdf](#)]

## References

1. Evans-Brown M, McVeigh J, Perkins C, Bellis M. Human Enhancement Drugs: The Emerging Challenges to Public Health. North West Public Health Observatory: Liverpool; Apr 11, 2012.
2. Liang B, Mackey T. Sexual medicine: Online risks to health--the problem of counterfeit drugs. *Nat Rev Urol* 2012 Sep;9(9):480-482. [doi: [10.1038/nrurol.2012.148](#)] [Medline: [22824776](#)]
3. Gray N. The evolution of online pharmacies. *Self Care Journal* 2011;2:76-86.
4. Ivanitskaya L, Brookins-Fisher J, Vibbert D, Erofeev D, Fulton L. Dirt cheap and without prescription: how susceptible are young US consumers to purchasing drugs from rogue internet pharmacies? *J Med Internet Res* 2010;12(2):e11 [FREE Full text] [doi: [10.2196/jmir.1520](#)] [Medline: [20439253](#)]
5. Maxwell SRJ, Webb D. Internet pharmacy: a web of mistrust? *Br J Clin Pharmacol* 2008 Aug;66(2):196-198 [FREE Full text] [doi: [10.1111/j.1365-2125.2008.03215.x](#)] [Medline: [18513280](#)]
6. Pal S, Laszlo K, Andras F, Gabriel H, Hajnal F, Adriana C, et al. Attitude Of Patients And Customers Regarding Purchasing Drugs Online. *Farmacia* 2015 Jan 01;63(1):93-98.
7. Fung C, Wooand H, Asch S. Controversies And Legal Issues Of Prescribing And Dispensing Medicines Using The Internet. *Mayo Clinics Proceedings* 2004;79:188-194.
8. Anand A, Sethil N, Sharon G. Internet Pharmacy: Need To Be Implemented In India. *Chronicles of Young Scientists* 2010;1(1):1-10.
9. Government Canada. 2003. Buying drugs over the Internet: its your health URL: <http://http://www.hc-sc.gc.ca/english/~dfliyhhuymn as internet.pdf> [accessed 2016-07-13] [WebCite Cache ID [6iyKIzFeP](#)]
10. Siegel-Itzkovich J. WHO calls for tighter controls on Internet. *BMJ* 1997 May 24;314(7093):1501-1501. [doi: [10.1136/bmj.314.7093.1501h](#)]
11. Mackey T, Liang B. The global counterfeit drug trade: patient safety and public health risks. *J Pharm Sci* 2011 Nov;100(11):4571-4579. [doi: [10.1002/jps.22679](#)] [Medline: [21698604](#)]
12. Corns C, Metcalfe K. Risks associated with herbal slimming remedies. *J R Soc Promot Health* 2002 Dec;122(4):213-219. [Medline: [12557729](#)]
13. Venhuis B, Van den Berg JD, Wagenaar H, Van RA, Barends D. Trends in drug substances detected in illegal weight-loss medicines and dietary supplements. *RIVM Report* 2009;370030002.
14. Chan K. Some aspects of toxic contaminants in herbal medicines. *Chemosphere* 2003 Sep;52(9):1361-1371. [doi: [10.1016/S0045-6535\(03\)00471-5](#)] [Medline: [12867165](#)]
15. World Health Organization. 2006. Counterfeit medicines: an update on estimates URL: <http://www.who.int/medicines/services/counterfeit/impact/TheNewEstimatesCounterfeit.pdf> [accessed 2016-07-13] [WebCite Cache ID [6iyLjHEJV](#)]
16. Giuliani PL. Interim Findings. 2004. Examination and Assessment of Prescription Drug Importation from Foreign Sources to the United States URL: <http://www.heartlanddorglndf/14978.ndf> [accessed 2016-07-13] [WebCite Cache ID [6iyLkfJ8q](#)]
17. Liang B, Mackey T. Searching for safety: addressing search engine, website, and provider accountability for illicit online drug sales. *Am J Law Med* 2009;35(1):125-184. [Medline: [19534258](#)]
18. Ancuceanu R, Dinu M, Arama C. Weight Loss Food Supplements: Adulteration And Multiple Quality Issues In Two Products Of Chinese Origin. *Farmacia* 2013;61:28-44.
19. Food and Drug Administration. 2010. The possible dangers of buying medicines over the Internet URL: <http://www.fda.gov/ForConsumers/ConsumerUpdates/ucm048396.htm> [accessed 2016-07-13] [WebCite Cache ID [6iyM3QFII](#)]
20. Brachmann S. IPWatchdog. 2015. 97 percent on Internet pharmacies pose a public threat URL: <http://www.ipwatchdog.com/2015/03/23/97-percent-of-internet-pharmacies-pose-a-public-safety-threat/id=55706/> [accessed 2016-07-13] [WebCite Cache ID [6iyM8fYSW](#)]
21. Interpol. 2015 Jun 16. Operations URL: <http://www.interpol.int/Crime-areas/Pharmaceutical-crime/Operations/Operation-Pangea> [accessed 2016-07-13] [WebCite Cache ID [6iyMCfnnQ](#)]
22. The One Show. 2014. £8.6 Million Drugs Seizure on BBC1 URL: <http://www.tinyurl.com/drugs-seizure> [accessed 2016-07-13] [WebCite Cache ID [6iyMHdapO](#)]
23. Rudolf PM, Bernstein IBG. Counterfeit drugs. *N Engl J Med* 2004 Apr 1;350(14):1384-1386. [doi: [10.1056/NEJMp038231](#)] [Medline: [15070787](#)]
24. Dooley K. Bloomberg News. 2004 Jun 17. Online pharmacies shipped fake drugs to US report says URL: <http://the.honoluluadvertiser.com/article/2004/Jun/17/bz/bz10a.html> [accessed 2016-07-13] [WebCite Cache ID [6iyMWfcQ5](#)]



25. Murphy T, Slotton D, Irvine K, Sukontason K, Goldman C. Mercury Contamination of Skin Whiteners in Cambodia. *Human and Ecological Risk Assessment: An International Journal* 2009 Nov 19;15(6):1286-1303. [doi: [10.1080/10807030903306877](https://doi.org/10.1080/10807030903306877)]
26. Zhang J, Wider B, Shang H, Li X, Ernst E. Quality of herbal medicines: challenges and solutions. *Complement Ther Med* 2012;20(1-2):100-106. [doi: [10.1016/j.ctim.2011.09.004](https://doi.org/10.1016/j.ctim.2011.09.004)] [Medline: [22305255](https://pubmed.ncbi.nlm.nih.gov/22305255/)]
27. Singh S, Prasad B, Savaliya A, Shah R, Gohil V, Kaur A. Strategies for characterizing sildenafil, vardenafil, tadalafil and their analogues in herbal dietary supplements, and detecting counterfeit products containing these drugs. *TrAC Trends in Analytical Chemistry* 2009 Jan 31;28(1):13-28. [doi: [10.1016/j.trac.2008.09.004](https://doi.org/10.1016/j.trac.2008.09.004)]
28. Shahverdi S, Hajimiri M, Pourmalek F, Torkamandi H, Gholami K, Hanafi S, et al. Iranian pharmacists' knowledge, attitude and practice regarding counterfeit drugs. *Iran J Pharm Res* 2012;11(3):963-968 [FREE Full text] [Medline: [24250525](https://pubmed.ncbi.nlm.nih.gov/24250525/)]
29. Desai K, Chewing B, Mott D. Health care use amongst online buyers of medications and vitamins. *Res Social Adm Pharm* 2015;11(6):844-858. [doi: [10.1016/j.sapharm.2015.01.001](https://doi.org/10.1016/j.sapharm.2015.01.001)] [Medline: [25769500](https://pubmed.ncbi.nlm.nih.gov/25769500/)]
30. BITLAW. 1946. Lanham Act URL: <http://www.bitlaw.com/source/15usc/1051.html> [accessed 2016-07-13] [WebCite Cache ID 6iyNRNpJa]
31. World Health Organization. Spurious/false-labelled/falsified/counterfeit (SFFC) medicines. 2016. URL: <http://www.who.int/mediacentre/factsheets/fs275/en/> [accessed 2016-07-13] [WebCite Cache ID 6iyNU3Gzd]
32. European Commission. 2015. Cosmetics URL: [http://ec.europa.eu/growth/sectors/cosmetics/index\\_en.htm](http://ec.europa.eu/growth/sectors/cosmetics/index_en.htm) [accessed 2016-07-13] [WebCite Cache ID 6iyNV78yE]
33. Spain J, Siegel C, Ramsey R. Selling drugs online: distribution - related legal/regulatory issues. *International Marketing Review* 2001 Aug;18(4):432-449. [doi: [10.1108/EUM0000000005935](https://doi.org/10.1108/EUM0000000005935)]
34. Henney JE. Cyberpharmacies and the role of the US Food And Drug Administration. *J Med Internet Res* 2001;3(1):E3 [FREE Full text] [doi: [10.2196/jmir.3.1.e3](https://doi.org/10.2196/jmir.3.1.e3)] [Medline: [11720945](https://pubmed.ncbi.nlm.nih.gov/11720945/)]
35. Andreassen H, Bujnowska-Fedak M, Chronaki C, Dumitru R, Pudule I, Santana S, et al. European citizens' use of E-health services: a study of seven countries. *BMC Public Health* 2007;7:53 [FREE Full text] [doi: [10.1186/1471-2458-7-53](https://doi.org/10.1186/1471-2458-7-53)] [Medline: [17425798](https://pubmed.ncbi.nlm.nih.gov/17425798/)]
36. Gray N, Klein J, Cantrill J. Buying medicines on the Internet: young consumers' perceptions. *Journal of the American Pharmaceutical Association* 2002;42:3-13.

## Abbreviations

- API:** active pharmaceutical ingredient  
**FDA:** Food and Drug Administration  
**MHRA:** Medicine and Healthcare Regulatory Agency  
**NHS:** National Health Service  
**PIL:** patient information leaflet

*Edited by G Eysenbach; submitted 27.11.15; peer-reviewed by A Al Hamid, T Mackey; comments to author 03.04.16; revised version received 21.05.16; accepted 10.06.16; published 18.07.16.*

*Please cite as:*

Assi S, Thomas J, Haffar M, Osselton D

*Exploring Consumer and Patient Knowledge, Behavior, and Attitude Toward Medicinal and Lifestyle Products Purchased From the Internet: A Web-Based Survey*

*JMIR Public Health Surveill* 2016;2(2):e34

URL: <http://publichealth.jmir.org/2016/2/e34/>

doi: [10.2196/publichealth.5390](https://doi.org/10.2196/publichealth.5390)

PMID: [27430264](https://pubmed.ncbi.nlm.nih.gov/27430264/)

©Sulaf Assi, Jordan Thomas, Mohamed Haffar, David Osselton. Originally published in *JMIR Public Health and Surveillance* (<http://publichealth.jmir.org>), 18.07.2016. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in *JMIR Public Health and Surveillance*, is properly cited. The complete bibliographic information, a link to the original publication on <http://publichealth.jmir.org>, as well as this copyright and license information must be included.

Protocol

# The Prevalence of Sexual Behavior Stigma Affecting Gay Men and Other Men Who Have Sex with Men Across Sub-Saharan Africa and in the United States

Shauna Stahlman<sup>1</sup>, MPH, PhD; Travis Howard Sanchez<sup>2</sup>, MPH, DVM; Patrick Sean Sullivan<sup>2</sup>, PhD, DVM; Sosthenes Ketende<sup>1</sup>, MSc; Carrie Lyons<sup>1</sup>, MPH; Manhattan E Charurat<sup>3</sup>, PhD; Fatou Maria Drame<sup>4,5</sup>, PhD; Daouda Diouf<sup>4</sup>, MSc; Rebecca Ezouatchi<sup>6</sup>, MA; Seni Kouanda<sup>7,8</sup>; Simplicie Anato<sup>9</sup>, MA; Tampose Mothopeng<sup>10</sup>; Zandile Mnisi<sup>11</sup>, BNSc; Stefan David Baral<sup>1</sup>, MPH, MD

<sup>1</sup>Center for Public Health and Human Rights, Department of Epidemiology, Johns Hopkins University, Baltimore, MD, United States

<sup>2</sup>Department of Epidemiology, Rollins School of Public Health, Emory University, Atlanta, GA, United States

<sup>3</sup>Division of Epidemiology and Prevention, Institute of Human Virology, University of Maryland, Baltimore, MD, United States

<sup>4</sup>Enda Santé, Dakar, Senegal

<sup>5</sup>Department of Geography, School of Social Sciences, Gaston Berger University, Saint-Louis, Senegal

<sup>6</sup>Enda Santé, Abidjan, Côte d'Ivoire

<sup>7</sup>Institut de Recherche en Sciences de la Santé-IRSS, Ouagadougou, Burkina Faso

<sup>8</sup>Institut Africain de Santé Publique, Ouagadougou, Burkina Faso

<sup>9</sup>Arc-en-ciel, Lomé, Togo

<sup>10</sup>Matrix Support Group, Maseru, Lesotho

<sup>11</sup>Ministry of Health, Mbabane, Swaziland

**Corresponding Author:**

Shauna Stahlman, MPH, PhD

Center for Public Health and Human Rights

Department of Epidemiology

Johns Hopkins University

615 N. Wolfe Street

Baltimore, MD, 21205

United States

Phone: 1 443 287 2370

Fax: 1 410 614 8371

Email: [sstahlm1@jhu.edu](mailto:sstahlm1@jhu.edu)

## Abstract

**Background:** There has been increased attention for the need to reduce stigma related to sexual behaviors among gay men and other men who have sex with men (MSM) as part of comprehensive human immunodeficiency virus (HIV) prevention and treatment programming. However, most studies focused on measuring and mitigating stigma have been in high-income settings, challenging the ability to characterize the transferability of these findings because of lack of consistent metrics across settings.

**Objective:** The objective of these analyses is to describe the prevalence of sexual behavior stigma in the United States, and to compare the prevalence of sexual behavior stigma between MSM in Southern and Western Africa and in the United States using consistent metrics.

**Methods:** The same 13 sexual behavior stigma items were administered in face-to-face interviews to 4285 MSM recruited in multiple studies from 2013 to 2016 from 7 Sub-Saharan African countries and to 2590 MSM from the 2015 American Men's Internet Survey (AMIS), an anonymous Web-based behavioral survey. We limited the study sample to men who reported anal sex with a man at least once in the past 12 months and men who were aged 18 years and older. Unadjusted and adjusted prevalence ratios were used to compare the prevalence of stigma between groups.

**Results:** Within the United States, prevalence of sexual behavior stigma did not vary substantially by race/ethnicity or geographic region except in a few instances. Feeling afraid to seek health care, avoiding health care, feeling like police refused to protect, being blackmailed, and being raped were more commonly reported in rural versus urban settings in the United States ( $P < .05$  for

all). In the United States, West Africa, and Southern Africa, MSM reported verbal harassment as the most common form of stigma. Disclosure of same-sex practices to family members increased prevalence of reported stigma from family members within all geographic settings ( $P < .001$  for all). After adjusting for potential confounders and nesting of participants within countries, AMIS-2015 participants reported a higher prevalence of family exclusion ( $P = .02$ ) and poor health care treatment ( $P = .009$ ) as compared with participants in West Africa. However, participants in both West Africa ( $P < .001$ ) and Southern Africa ( $P < .001$ ) reported a higher prevalence of blackmail. The prevalence of all other types of stigma was not found to be statistically significantly different across settings.

**Conclusions:** The prevalence of sexual behavior stigma among MSM in the United States appears to have a high absolute burden and similar pattern as the same forms of stigma reported by MSM in Sub-Saharan Africa, although results may be influenced by differences in sampling methodology across regions. The disproportionate burden of HIV is consistent among MSM across Sub-Saharan Africa and the United States, suggesting the need in all contexts for stigma mitigation interventions to optimize existing evidence-based and human-rights affirming HIV prevention and treatment interventions.

(*JMIR Public Health Surveill* 2016;2(2):e35) doi:[10.2196/publichealth.5824](https://doi.org/10.2196/publichealth.5824)

## KEYWORDS

stigmatization; social stigma; HIV; male homosexuality; United States; Western Africa; Southern Africa; mental health

## Introduction

The World Health Organization, UNAIDS, and the White House National HIV/AIDS Strategy have collectively called for the need to reduce stigma toward key populations to reduce the impact of human immunodeficiency virus (HIV) and to optimize HIV treatment outcomes [1-3]. In this context, gay men and other men who have sex with men (MSM) continue to be among the populations at highest risk for HIV infection worldwide [4,5]. The Centers for Disease Control and Prevention recently estimated that if current incidence rates continue, 1 in 6 MSM in the United States will be diagnosed with HIV in their lifetime, including 1 in 2 Black MSM, 1 in 4 Latino MSM, and 1 in 11 White MSM [6]. In low- and middle-income countries, MSM are estimated to have at least 19 times the odds of living with HIV compared with other reproductive-aged adults [7]. And in Sub-Saharan Africa, even in regions commonly characterized as generalized epidemics, the prevalence and incidence of HIV among MSM is consistently higher than that of age-matched men in all settings [8,9].

Sexual behavior stigma, defined here as stigma that is anticipated, perceived, or experienced as a result of one's sexual experience [10], has been linked with adverse HIV-related outcomes among MSM including reduced rates of HIV testing and increased sexual risk practices leading to enhanced risk for HIV infection [11-14]. Separately, sexual behavior stigma has been linked to adverse mental health outcomes such as depression, suicidal ideation, and substance use disorders [15-17]. This stigma can operate at community- or structural-levels to negatively influence health outcomes. At the community level, sexual behavior-related stigma can limit the provision and uptake of sexual health services [18,19]. For example, culturally insensitive health workers may result in MSM avoiding HIV prevention services; or even more problematically, MSM living with HIV may avoid HIV treatment. Reduced use of health and HIV services by MSM, due to enacted or perceived discrimination, may limit knowledge of the risks of unprotected anal intercourse and opportunities for access to prevention services [20,21]. At the structural level, laws against same-sex practices can increase fear and avoidance

of health care services, risk behaviors, stress, and promote violence, which can worsen health conditions for MSM and the broader communities [20,22-25].

Within the United States, HIV disproportionately affects young, Black and Hispanic MSM [26-32]. However, individual behavioral risk factors do not appear to explain the increased risk for HIV among racial/ethnic minorities [30,33]. Instead, the burden of stigma and mental health secondary to sexual and gender minority stress may contribute to the disparities observed around the world [34,35]. As a result, there is a growing recognition for the need to be able to measure and evaluate stigma toward MSM.

However, a consensus is lacking for how to quantify, measure, and evaluate sexual behavior stigma for key populations. Previous studies have identified high levels of sexual behavior stigma among MSM both in Sub-Saharan Africa and in the United States including physical assault and verbal harassment [23,36-39]. However, comparisons are imperfect because many studies do not use consistent metrics for measuring this stigma across settings [40]. Thus, the objectives of this paper are (1) to describe the prevalence of these sexual behavior stigma items in the United States and across age, racial/ethnic, population density, and regional subgroups, and (2) to make a comparison of sexual behavior stigma between MSM in 3 different global regions (United States, West Africa, and Southern Africa) using consistent metrics.

## Methods

### Study Population and Key Measures

#### *American Men's Internet Survey (AMIS)*

The American Men's Internet Survey (AMIS) is an annual cross-sectional behavioral survey of MSM in the United States [41,42]. AMIS-2015 recruited MSM from a variety of websites using banner advertisements or email blasts. Websites included general social networking sites (eg, Facebook), general gay interest sites (eg, sites that post news and articles relevant to a gay audience), gay social networking sites, and mobile-only geospatial social networking applications that connect men with

other men based on their proximity. Approximately one-half of the surveys were performed on desktop computers and one-half on mobile phones. We used an interim survey dataset from September 2015 to March 2016. To be eligible to participate, men had to be aged 15 years or older, identify as the male gender, and report that they had oral or anal sex with a man at least once in the past. For the purpose of making comparisons with the African data sets, we limited the study sample to men who reported anal sex with a man at least once in the past 12 months and men who were aged 18 years and older, because these were eligibility requirements in the African studies. We also limited the AMIS-2015 sample to men with a current US residency as measured using reported zip codes and Internet protocol addresses. No incentive was provided to the participants.

AMIS consisted of a core questionnaire that was administered to all participants and included questions about demographics and disclosure status, which refers to whether the participant disclosed his same sex practices to either health care workers or family members. It also consisted of 3 different subset questionnaires to which participants were randomized at the start of the survey. All AMIS-2015 participants were asked if they identified as Hispanic/Latino, American Indian or Alaskan Native, Asian, Black or African American, Native Hawaiian or Pacific Islander, or White, and could choose all groups that apply. For the purpose of these analyses, we created a single variable for race/ethnicity, which grouped participants into the following mutually exclusive categories: Hispanic, non-Hispanic Black (“Black”), non-Hispanic White (“White”), and non-Hispanic Other (“Other”), which included those who identified as multiracial, Asian, and Native American/Hawaiian. Participants who indicated that they preferred not to answer questions about race, or that they didn’t know/did not apply were treated as missing (39/2551, 1.5%). We used a combination of county and zip code of residence to determine population

density. Population density was classified as urban if participants were recruited from a county or zip code that had a population density of at least 1000 people per square mile, and rural if less than 1000 people per square mile. Sexual behavior stigma items were included in a subset questionnaire; therefore, one-third of participants were asked about sexual behavior stigma (2590/7853, 33.0%).

### Sub-Saharan Africa

Data from Sub-Saharan Africa have been presented in previous studies [23,43-45]. In this secondary analysis, we created a combined data set that was limited to MSM who were aged 18 years or older, reported being assigned male sex at birth and also identified as male, and reported anal intercourse with a man in the past 12 months. Data were collected predominantly via respondent driven sampling (RDS) [46] and snowball sampling [47] from 4285 MSM from 7 Sub-Saharan African countries (Table 1). RDS seeds were recruited from local MSM-affiliated community-based organizations (CBOs) although not all seeds were CBO members. Seeds were selected to represent the diversity of the target population with respect to demographics, HIV status, partnering, and other factors. There is significant consistency of methods across settings with face-to-face administration of a structured survey instrument by MSM or MSM-friendly staff members at the local CBOs. In addition, all data sets contain measures of sociodemographics, disclosure status, and sexual behavior stigma. For these analyses, population density was similarly classified as urban if participants were recruited from a study site located in a city that had a population density of at least 1000 people per square mile, and rural if less than 1000 people per square mile. Participants received a modest reimbursement for their time, the cost of travel to the study site (US\$2-\$6), and for each eligible participant they recruited into the study (for RDS studies).

**Table 1.** Summary of Sub-Saharan Africa MSM data sets.

Region	Country	Laws pertaining to same sex practices [48-50]	Recruitment dates	Study site locations	Recruitment method	Sample size
West Africa	Burkina Faso	Not criminalized	January-August 2013	Bobo Dioulasso and Ouagadougou	RDS <sup>a</sup>	443
West Africa	Cote d’Ivoire	Not criminalized	March-October 2015	Abidjan, Gagnoa, Bouaké, and Yamoussoukro	RDS	794
Southern Africa	Lesotho	Sodomy prohibited as common-law offence	February- September 2014	Maseru and Maputsoe	RDS	487
West Africa	Nigeria	Imprisonment	March 2013- August 2015	Abuja and Lagos	RDS	1067
West Africa	Senegal	Imprisonment	February- November 2015	Dakar, Mbour, and Thies	RDS	522
Southern Africa	Swaziland	Unenforced penalties	October- December 2014	Multiple	Snowball	419
West Africa	Togo	Imprisonment	January-June 2013	Kara and Lome	RDS	553

<sup>a</sup>Abb: respondent driven sampling.



## Sexual Behavior Stigma

Sexual behavior stigma was measured using the same 13 items in AMIS-2015 and in each of the Sub-Saharan Africa studies (Textbox 1). These measures include perceived, anticipated, and experienced stigma, such as stigma from family and friends,

stigma from health care workers, and stigma from broader society. For questions pertaining to physical attacks and rape, we included responses as a “yes” only if the participant believed that these experiences were related to the fact that he has sex with men.

**Textbox 1.** Sexual behavior stigma items (response options: yes/no).

1. Have you ever felt excluded from family activities because you have sex with men?
2. Have you ever felt that family members have made discriminatory remarks or gossiped about you because you have sex with men?
3. Have you ever felt rejected by your friends because you have sex with men?
4. Have you ever felt afraid to go to health care services because you worry someone may learn you have sex with men?
5. Have you ever avoided going to health care services because you worry someone may learn you have sex with men?
6. Have you ever felt that you were not treated well in a health center because someone knew that you have sex with men?
7. Have you ever heard health care providers gossiping about you (talking about you) because you have sex with men?
8. Have you ever felt that the police refused to protect you because you have sex with men?
9. Have you ever felt scared to be in public places because you have sex with men?
10. Have you ever been verbally harassed and felt it was because you have sex with men?
11. Have you ever been blackmailed by someone because you have sex with men?
- 12a. Has someone ever physically hurt you (pushed, shoved, slapped, hit, kicked, choked or otherwise physically hurt you)?
- 12b. Do you believe any of these experiences of physical violence was/were related to the fact that you have sex with men?
- 13a. Have you ever been forced to have sex when you did not want to? (By forced, I mean physically forced, coerced to have sex, or penetrated with an object, when you did not want to).
- 13b. Do you believe any of these experiences of sexual violence were related to the fact that you have sex with men?

## Ethics

AMIS-2015 data collection was approved by the human subjects research review board at Emory University. Sub-Saharan African studies were approved by respective in-country ethics committees: the Health Research Ethics Committee of Burkina Faso, the Health Research Ethics Committee of Côte d'Ivoire, the Lesotho National Health Research Ethics Committee, the Ethical Committee of Togo, the Senegalese National Health Research Ethics, the Swaziland Scientific Ethics Committee, and the institutional review board (IRB) at the Johns Hopkins Bloomberg School of Public Health. For Nigeria, approval was obtained by the Federal Capital Territory Health Research Ethics Committee, the University of Maryland Baltimore IRB, and the Walter Reed Army Institute of Research IRB. Informed consent was obtained from all individual participants included in these analyses.

## Statistical Analysis

We used descriptive statistics (frequencies, percentages) to describe the distribution of sexual behavior stigma items between and within groups. In bivariate analyses, we used log-binomial regression models to generate crude prevalence ratios (PRs) including 95% confidence intervals, which were used to test variables of interest for associations with stigma. Specifically, we examined the association of race/ethnicity, US region, population density, and age group with each sexual behavior stigma item in AMIS-2015. Within US/Africa regions, we measured the association between disclosure and stigma from family members and health care workers.

Multivariable log-binomial models were used to test for associations between region (United States, West Africa, Southern Africa) and prevalence of each stigma item after adjusting for potential confounders (age, disclosure, population density, and education level), which were identified based on findings from previous studies [16,39,43,51,52]. In addition, we accounted for nesting of participants within countries using a random intercept in each of the multivariable models, which is a method that has been used in similar studies [20,53,54]. Listwise deletion was used to handle missing data. Significance was determined at  $\alpha = .05$ .

## Results

### Sample Demographics

In AMIS-2015, the median age of participants was 32 years (interquartile range (IQR)=24-50) (Table 2). The sample was roughly evenly divided across regions of the United States, with 19.9% (516/2589) from the Midwest, 18.3% (474/2589) from the Northeast, 37.1% (961/2589) from the South, and 24.6% (638/2589) from the West. Most lived in an urban (1749/2551, 68.6%) area. The distribution of race/ethnicity was as follows: 13.4% (342/2551) Hispanic, 7.4% (188/2551) non-Hispanic Black, 71.7% (1829/2551) non-Hispanic White, and 7.5% (192/2551) Other. The majority (1368/2569, 53.3%) had completed a college level education or higher, 35.0% (898/2569) completed some college, and 11.8% (303/2569) completed high school or lower. Most participants had disclosed their same-sex practices to either a family member (1982/2412, 82.2%) or health care worker (1735/2397, 72.4%).



In West Africa, participants were recruited from Nigeria (1067/3379, 31.6%), Cote d'Ivoire (794/3379, 23.5%), Togo (553/3379, 16.4%), Senegal (522/3379, 15.5%), and Burkina Faso (443/3379, 13.1%). The median age was 23 years (IQR=21-27). Most had completed at least a secondary school education (1898/3359, 56.5%), whereas 24.8% (834/3359) completed more than a secondary school education and 18.7% (627/3359) completed primary school or lower. Overall, 19.7% (665/3379) had disclosed their same-sex practices to a family member and 35.6% (1199/3379) had disclosed to a health care worker.

In Southern Africa, participants were recruited from Lesotho (487/906, 53.8%) and Swaziland (419/906, 46.3%). The median age was 24 years (IQR=21-28). Similar to West Africa, most (597/905, 66.0%) had completed a secondary school education, whereas 20.8% (188/905) completed more than a secondary school education, and 13.3% (120/905) completed primary school or less. Just less than one-third (277/901, 30.7%) disclosed same-sex practices to a family member and 12.2% (110/901) disclosed to a health care worker.

**Table 2.** Sample demographics by United States/Africa region.

	United States N (%)	Southern Africa N (%)	West Africa N (%)
<b>Median age (IQR)</b>	32 (24-50)	24 (21-28)	23 (21-27)
<b>Education completed</b>			
High school/secondary school or less	303 (11.8)	717 (79.2)	2525 (75.2)
More than secondary/high school	2266 (88.2)	188 (20.8)	834 (24.8)
<b>Population density</b>			
Rural	802 (31.4)	603 (66.6)	597 (17.7)
Urban	1749 (68.6)	303 (33.4)	2782 (82.3)
<b>Disclosed same-sex practices to health care worker</b>	1735 (72.4)	110 (12.2)	1199 (35.6)
<b>Disclosed same-sex practices to family member</b>	1982 (82.2)	277 (30.7)	665 (19.7)

### Sexual Behavior Stigma in AMIS-2015

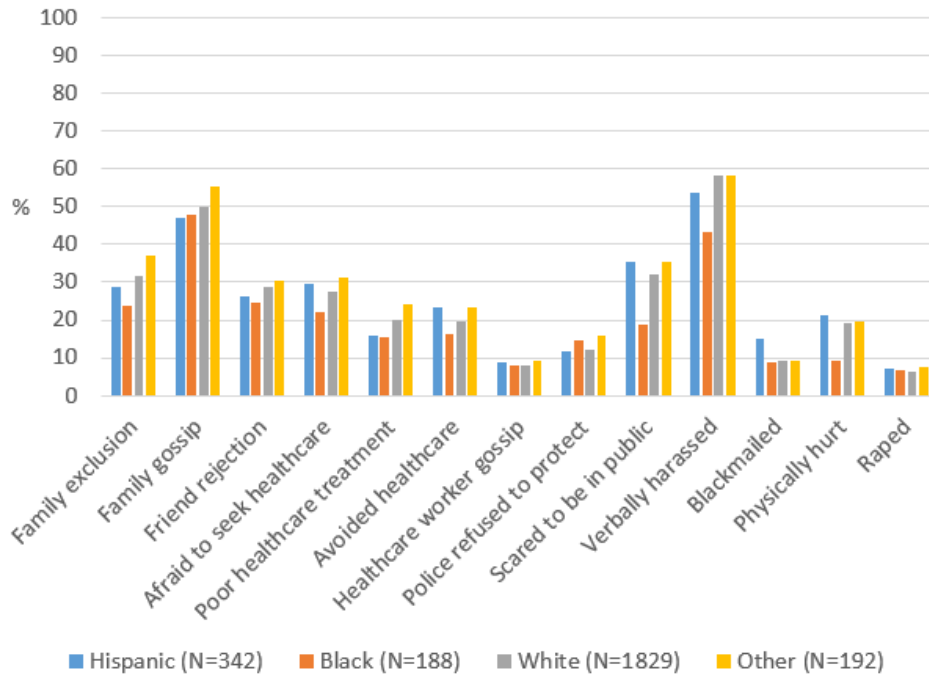
Overall prevalence of stigma items in AMIS-2015 are presented in [Table 3](#). Briefly, the most commonly reported items were verbal harassment (1423/2509, 56.7%), family gossip (1155/2312, 50.0%), and being scared to be in public (811/2551, 31.8%). Prevalence of sexual behavior stigma by race/ethnicity within the United States are shown in [Figure 1](#) and [Multimedia Appendix 1](#). In the bivariate analyses, Hispanic MSM were more likely to report being blackmailed as compared with White MSM ( $P=.002$ ). Black MSM were less likely than White MSM to report family exclusion ( $P=.04$ ), fear of being in public ( $P<.001$ ), verbal harassment ( $P<.001$ ), and physical attacks ( $P=.002$ ).

There were no differences in sexual behavior stigma prevalence by United States region, except that MSM in the West were less likely to feel like police refused to protect them as compared with MSM in the South ( $P=.04$ ), and MSM in the Midwest

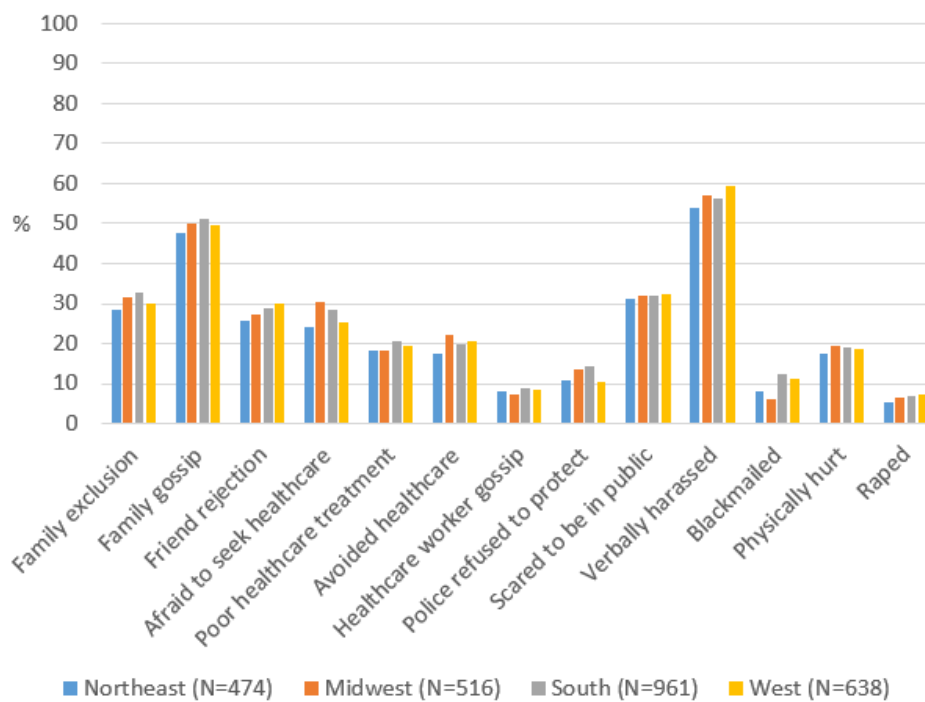
( $P<.001$ ), and Northeast ( $P=.02$ ) were less likely to report blackmail as compared with MSM in the South ([Figure 2](#) and [Multimedia Appendix 2](#)). MSM who lived in rural as compared with urban areas were more likely to report being afraid to seek health care ( $P<.001$ ), avoiding health care ( $P=.003$ ), feeling like police refused to protect ( $P=.001$ ), being blackmailed ( $P=.004$ ), and being raped ( $P=.04$ ) ([Figure 3](#) and [Multimedia Appendix 3](#)).

Finally, young MSM (aged 18-24 years) were more likely than older MSM to be excluded by family members ( $P=.002$ ), gossiped about by family members ( $P=.007$ ), afraid to seek health care ( $P<.001$ ), avoid seeking health care ( $P=.002$ ), scared to be seen in public ( $P=.03$ ), and to be blackmailed ( $P=.003$ ) ([Figure 4](#) and [Multimedia Appendix 4](#)). However, older MSM (aged 25 years and older) were more likely to have been treated poorly in a health care center ( $P<.001$ ), and to feel like police refused to protect ( $P=.03$ ).

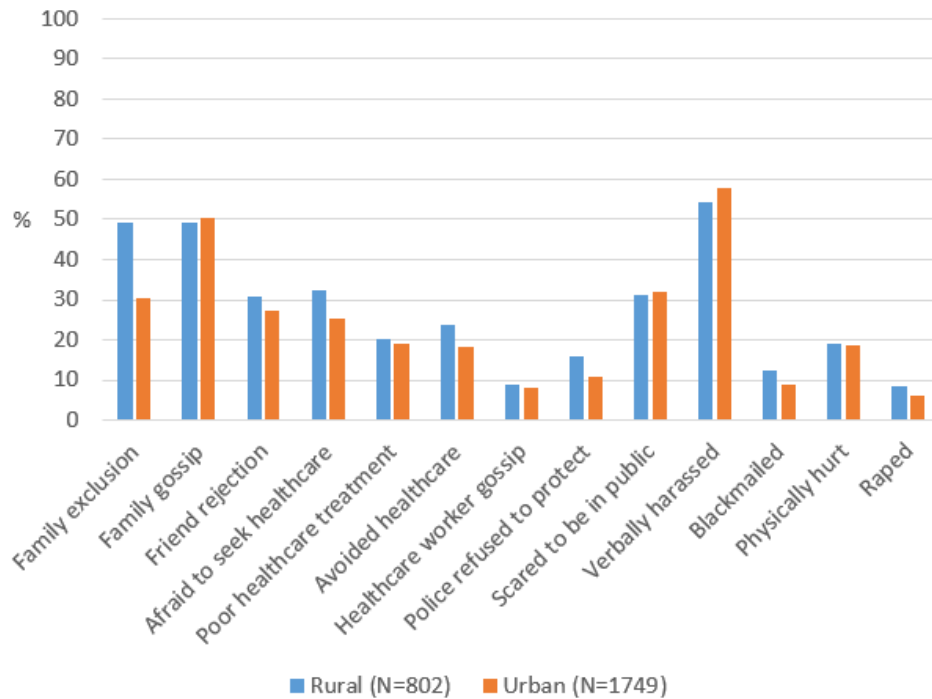
**Figure 1.** Prevalence of sexual behavior stigma among MSM in AMIS-2015 by race/ethnicity.



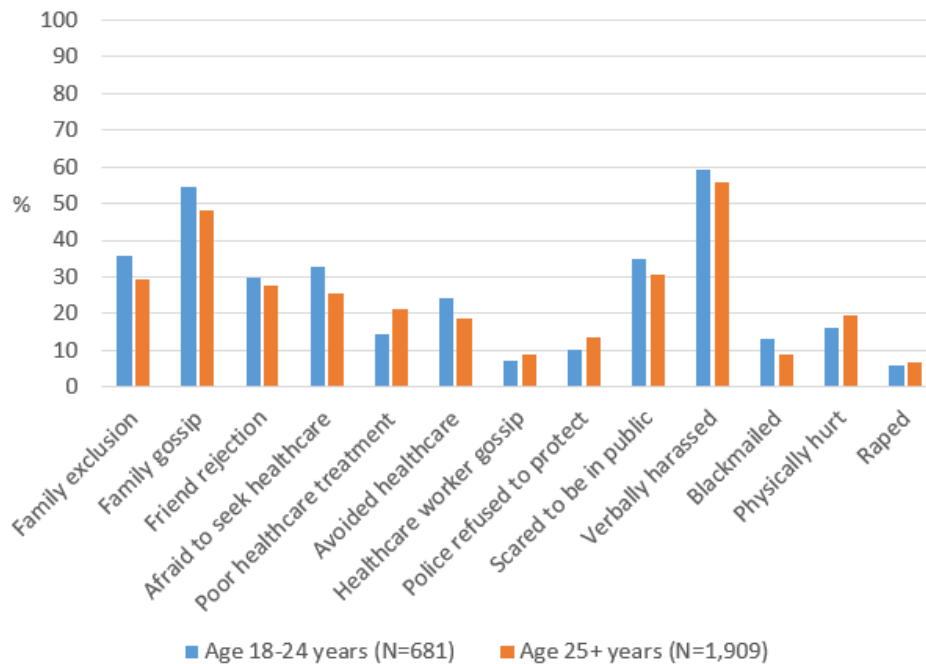
**Figure 2.** Prevalence of sexual behavior stigma among MSM in AMIS-2015 by United States region.



**Figure 3.** Prevalence of sexual behavior stigma among MSM in AMIS-2015 by population density.



**Figure 4.** Prevalence of sexual behavior stigma among MSM in AMIS-2015 by age group.



**Comparison of Stigma Prevalence Across United States and Sub-Saharan Africa**

Overall, sexual behavior stigma reported in AMIS-2015 was similar to or higher than what was reported by participants in the Sub-Saharan African studies (Figure 5 and Table 3). Family exclusion, family gossip, friend rejection, being afraid to seek health care, being treated poorly by a health care worker, feeling like police refused to protect, being scared to be in public, verbal harassment, and being physically hurt were all more commonly reported by AMIS-2015 participants ( $P < .001$  for all, except

$P = .005$  for comparing fear of seeking health care with Southern Africa). AMIS-2015 participants were also more likely than MSM in West Africa to report avoiding seeking health care ( $P = .002$ ) and health care worker gossip ( $P < .001$ ). Participants in West and Southern Africa more commonly reported blackmail ( $P < .001$  for both) compared with those in the United States, and participants in West Africa more commonly reported rape as compared with MSM in the United States ( $P < .001$ ). In all settings, MSM reported verbal harassment as the most common experience of stigma. Rape was the least common experience in United States and Southern Africa, whereas being treated

poorly in a health care center was least commonly reported in West Africa.

**Table 3.** Prevalence of sexual behavior stigma among MSM by United States/Africa region.

Stigma	Region	n/N (%)	Prevalence ratio (95% confidence interval)	P value
<b>Family exclusion</b>	United States	775/2489 (31.1)	Reference	--
	Southern Africa	126/905 (13.9)	0.45 (0.38-0.53)	<.001
	West Africa	298/3378 (8.8)	0.28 (0.25-0.32)	<.001
<b>Family gossip</b>	United States	1155/2312 (50.0)	Reference	--
	Southern Africa	181/903 (20.0)	0.40 (0.35-0.46)	<.001
	West Africa	727/3377 (21.5)	0.43 (0.40-0.47)	<.001
<b>Friend rejection</b>	United States	677/2400 (28.2)	Reference	--
	Southern Africa	173/902 (19.2)	0.68 (0.59-0.79)	<.001
	West Africa	604/3378 (17.9)	0.63 (0.58-0.70)	<.001
<b>Afraid to seek health care</b>	United States	665/2429 (27.4)	Reference	--
	Southern Africa	203/905 (22.4)	0.82 (0.71-0.94)	.005
	West Africa	756/3378 (22.4)	0.82 (0.75-0.89)	<.001
<b>Poor healthcare treatment</b>	United States	462/2374 (19.5)	Reference	--
	Southern Africa	62/905 (6.9)	0.35 (0.27-0.45)	<.001
	West Africa	102/3355 (3.0)	0.16 (0.13-0.19)	<.001
<b>Avoided health care</b>	United States	488/2430 (20.1)	Reference	--
	Southern Africa	180/905 (19.9)	0.99 (0.85-1.15)	.90
	West Africa	573/3377 (17.0)	0.84 (0.76-0.94)	.002
<b>Health care worker gossip</b>	United States	201/2397 (8.4)	Reference	--
	Southern Africa	74/905 (8.2)	0.98 (0.76-1.26)	.85
	West Africa	195/3350 (5.8)	0.69 (0.57-0.84)	<.001
<b>Police refused to protect</b>	United States	298/2367 (12.6)	Reference	--
	Southern Africa	65/902 (7.2)	0.57 (0.44-0.74)	<.001
	West Africa	244/3374 (7.2)	0.57 (0.49-0.67)	<.001
<b>Scared to be in public</b>	United States	811/2551 (31.8)	Reference	--
	Southern Africa	208/905 (23.0)	0.72 (0.63-0.83)	<.001
	West Africa	452/3376 (13.4)	0.42 (0.38-0.47)	<.001
<b>Verbally harassed</b>	United States	1423/2509 (56.7)	Reference	--
	Southern Africa	351/905 (38.8)	0.68 (0.63-0.75)	<.001
	West Africa	934/3377 (27.7)	0.49 (0.46-0.52)	<.001
<b>Blackmailed</b>	United States	252/2507 (10.1)	Reference	--
	Southern Africa	178/905 (19.7)	1.96 (1.64-2.33)	<.001
	West Africa	686/3378 (20.3)	2.02 (1.77-2.31)	<.001
<b>Physically hurt</b>	United States	473/2513 (18.8)	Reference	--
	Southern Africa	120/903 (13.3)	0.71 (0.59-0.85)	<.001
	West Africa	388/3356 (11.6)	0.61 (0.54-0.70)	<.001
<b>Raped</b>	United States	159/2377 (6.7)	Reference	--
	Southern Africa	59/902 (6.5)	0.98 (0.73-1.31)	.88
	West Africa	388/3365 (11.5)	1.72 (1.44-2.06)	<.001

Sexual behavior stigma from family and friends was higher among those who had disclosed their sexual practices to a family member in all regions ( $P < .001$  for all); however, the prevalence of friend rejection among AMIS-2015 participants was similar between the two disclosure groups ( $P = .75$ ) (Figure 6 and Multimedia Appendix 5).

In the United States, participants who had disclosed their sexual practices to a health care worker were less likely to be afraid to

seek health care ( $P < .001$ ) or to avoid seeking health care ( $P < .001$ ) as compared with those who had not disclosed; however, they were more likely than those who had not disclosed to report being treated poorly in a health care center ( $P < .001$ ) (Figure 7 and Multimedia Appendix 6). In both West and Southern Africa, the prevalence of each health care-related sexual behavior stigma item increased among those who had disclosed versus not disclosed ( $P < .001$  for all).

Figure 5. Prevalence of sexual behavior stigma among MSM by United States/Africa region.

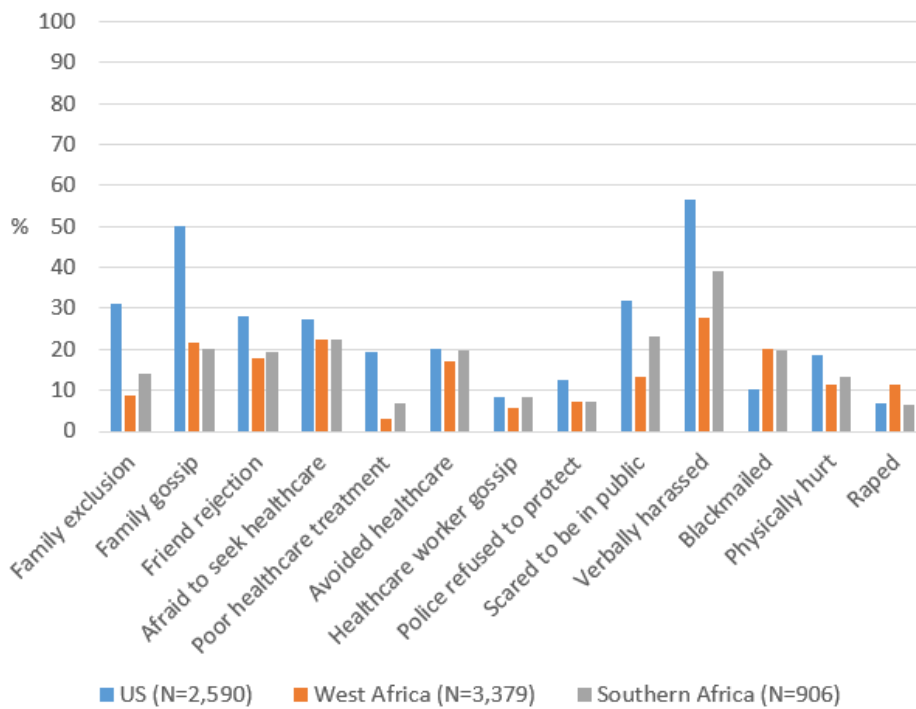
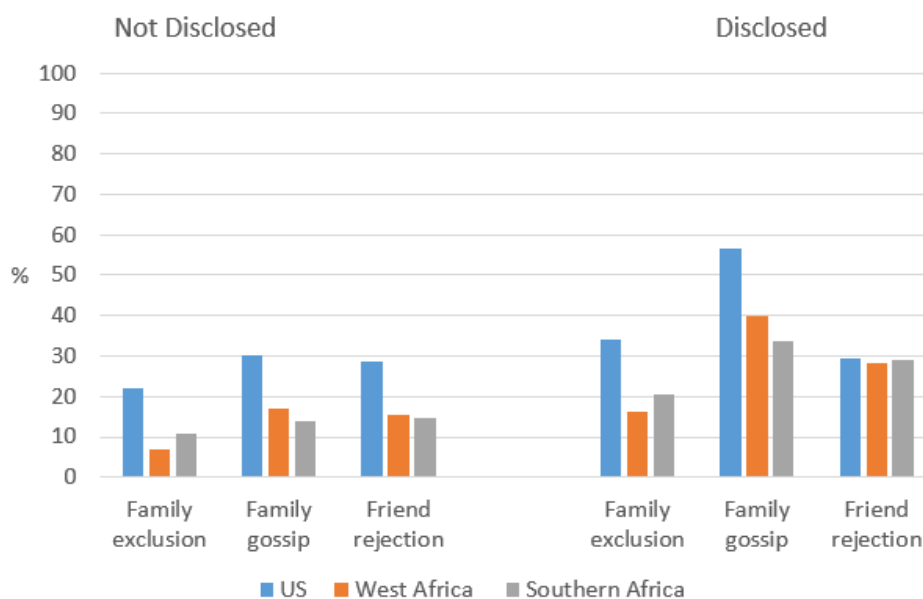
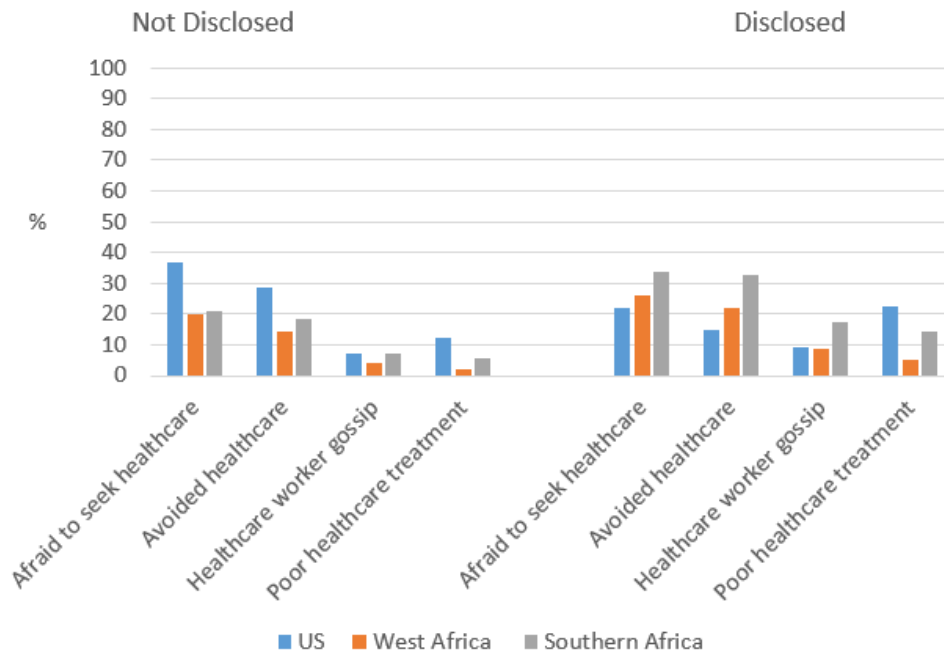


Figure 6. Prevalence of sexual behavior stigma among MSM who disclosed same-sex behaviors to family vs not disclosed.





**Figure 7.** Prevalence of sexual behavior stigma among MSM who disclosed same-sex behaviors to health care worker vs not disclosed.



**Multivariable Adjusted Associations of United States/Africa Region With Sexual Behavior Stigma Items**

After adjusting for age, disclosure status with family members and health care workers, education level, and population density, we found that AMIS-2015 participants continued to report

higher levels of family exclusion ( $P=.02$ ) and poor health care treatment ( $P=.009$ ) as compared with West Africa (Table 4). However, participants in both West Africa ( $P<.001$ ) and Southern Africa ( $P<.001$ ) reported a higher prevalence of blackmail. The prevalence of all other types of stigma was not found to be statistically significantly different across settings.

**Table 4.** Adjusted associations of United States/Africa region with sexual behavior stigma.

Stigma	Region	Adjusted prevalence ratio <sup>a</sup> (95% CI)	P value
<b>Family exclusion</b>	United States	Reference	--
	Southern Africa	0.52 (0.19-1.44)	.21
	West Africa	0.33 (0.13-0.82)	.02
<b>Family gossip</b>	United States	Reference	--
	Southern Africa	0.52 (0.20-1.35)	.18
	West Africa	0.50 (0.22-1.17)	.11
<b>Friend rejection</b>	United States	Reference	--
	Southern Africa	0.91 (0.42-1.97)	.81
	West Africa	0.75 (0.37-1.49)	.41
<b>Afraid to seek health care</b>	United States	Reference	--
	Southern Africa	0.79 (0.20-3.18)	.74
	West Africa	0.77 (0.22-2.67)	.68
<b>Poor health care treatment</b>	United States	Reference	--
	Southern Africa	0.52 (0.14-1.95)	.33
	West Africa	0.21 (0.06-0.67)	.009
<b>Avoided health care</b>	United States	Reference	--
	Southern Africa	0.91 (0.20-4.17)	.09
	West Africa	0.85 (0.22-3.32)	.82
<b>Health care worker gossip</b>	United States	Reference	--
	Southern Africa	1.25 (0.32-4.96)	.75
	West Africa	0.97 (0.29-3.30)	.97
<b>Police refused to protect</b>	United States	Reference	--
	Southern Africa	0.84 (0.08-8.38)	.88
	West Africa	0.59 (0.08-4.59)	.61
<b>Scared to be in public</b>	United States	Reference	--
	Southern Africa	0.71 (0.15-3.28)	.66
	West Africa	0.37 (0.09-1.45)	.15
<b>Verbally harassed</b>	United States	Reference	--
	Southern Africa	0.84 (0.44-1.59)	.59
	West Africa	0.52 (0.30-0.93)	.03
<b>Blackmailed</b>	United States	Reference	--
	Southern Africa	2.94 (1.70-5.07)	<.001
	West Africa	2.86 (1.76-4.64)	<.001
<b>Physically hurt</b>	United States	Reference	--
	Southern Africa	0.83 (0.17-4.07)	.92
	West Africa	0.56 (0.14-2.31)	.43
<b>Raped</b>	United States	Reference	--
	Southern Africa	0.94 (0.17-5.15)	.94
	West Africa	1.68 (0.37-7.62)	.50

<sup>a</sup>Models adjust for age, disclosure of same-sex behaviors to family or health care workers, education level, and population density. Models include a random intercept to account for nesting of participants within countries.

## Discussion

### Principal Results

In these analyses, we identified a high prevalence of experienced, anticipated, and perceived sexual behavior stigma for MSM in all settings. Based on [Figure 5](#), the overall pattern of responses appeared to be similar across regions, with MSM in all settings most commonly reporting verbal harassment and less commonly reporting rape or being treated poorly in a health care center. Surprisingly, in many cases US participants reported even greater levels of stigma than MSM in settings where same sex practices were criminalized. However, many of these differences were not significant after adjusting for potential confounders, and may also be the result of inherent differences in sampling methodology (eg, Web-based questionnaire vs face-to-face interviews).

### Interpretation and Comparison with Prior Work

Traumatic experiences among MSM are problematic for the HIV response because they can lead to reduced uptake and use of health services [18,20,21,23,55]. Indeed, stigma has been associated with reduced rates of HIV testing, increased risk for HIV infection, increased fear and avoidance of health care, increased condomless anal sex, and reduced engagement in HIV treatment for those living with HIV [11-14,23]. Stigma has further been linked to adverse mental health outcomes such as reduced self-esteem, internalized homophobia, and depression, with potential mediation by resiliency and coping ability [56-59]. Thus, the findings of high prevalence of multiple forms of stigma suggest the potential for stigma mitigation interventions to improve existing mental health services and HIV prevention and treatment interventions for MSM.

Within AMIS-2015, there were minimal differences in report of sexual behavior stigma by race/ethnicity and by region, suggesting the pervasiveness of these experiences or perceptions among MSM in the United States. However, MSM in rural areas were more likely to report fear and avoidance of health care, feeling like police refused to protect, being blackmailed, and being raped. This is consistent with the findings from previous studies suggesting that stigma toward accessing mental health and HIV services is higher in rural regions of both the United States and Sub-Saharan Africa as compared with more urban regions [51,60,61]. Our finding that Black MSM experienced lower levels of certain types of sexual behavior stigma is in contrast to previous literature suggesting that stigma surrounding sexual orientation is more pervasive in the Black community [62,63]. However, much of this previous work assessed internalized homo-negativity among Black MSM, which we did not measure in this study [64]. In addition, there is potential selection bias in that Black MSM are underrepresented in AMIS-2015, and thus this finding should be further explored and confirmed using alternative sampling methods. There were some differences within AMIS-2015 between age groups, with young MSM (aged 18-24 years) being more likely than older MSM to report stigma from family, to be afraid to seek health care or avoid seeking health care, to be scared to be seen in public, and to be blackmailed. This is somewhat surprising given that we measured lifetime exposure to stigma and

suspected that older MSM would have had more opportunity for exposure, and thus higher levels of stigma. These high levels of anticipated stigma and particularly fear and avoidance of health care among young MSM are important to address, as they likely contribute to the sustained or growing incidence of HIV among young MSM worldwide [65-69].

When we stratified by disclosure status, the prevalence of stigma from family members increased for participants who had disclosed same-sex practices to family as compared with those who had not disclosed, and this occurred within each setting of the United States, West Africa, and Southern Africa. This is likely reflective of the fact that in all of these settings MSM who have disclosed their same sex practices are more easily identified as targets for stigma, discrimination, and harassment [43,70-72]. However, the patterns of stigma associated with disclosure status were somewhat different when we examined health care-related stigma. In the United States, those who had disclosed same-sex practices to a health care worker were less likely to fear or avoid seeking health care services, although they were more likely to report being treated poorly in a health care center. Based on these findings, it seems possible that MSM who have disclosed to their providers are more comfortable seeking health care, even though it can sometimes result in negative experiences. In Sub-Saharan Africa, health care stigma increased among those who had disclosed versus not disclosed, indicating the immediate need for structural interventions to improve access to culturally competent health care for MSM across all settings [1,22,73].

There have been successful efforts in Sub-Saharan Africa and other regions to increase clinical and cultural competency for health care workers who provide HIV and sexually transmitted infection prevention, treatment, and care to MSM patients [74-78]. Given our findings, these efforts should be intensified and expanded to cover all domains of health care, particularly because the proper training of health care professionals tends to be one of the more easily implementable intervention strategies to reduce or mitigate stigma. However, stigma toward MSM is not limited to health care settings and structural interventions may also be needed to reduce stigma in community settings including schools, workplaces, churches, and families. These interventions would need to be appropriately tailored to meet the needs of different cultures and communities. In the United States, for example, gay-straight alliances have been successful in reducing homophobia in schools, although this method may or may not be suitable for African settings [79]. Eventual acceptance of lesbian, gay, bisexual, and transgender individuals may be inevitable with slowly increasing social acceptance over time; however, these social changes are not happening at the pace required to make an immediate impact on reducing negative health outcomes.

### Limitations

Although we restricted the study populations from each data set to be as similar as possible (eg, aged 18 years and older, anal sex with a man in the past 12 months, cis-gender male) and performed adjusted analyses, there are some inherent differences between the United States and Sub-Saharan Africa study populations that cannot be adjusted for, including sampling

strategies, mode of survey administration, and time period of study participation. Because AMIS-2015 was an anonymous Web-based survey, it is possible some participants were more comfortable to disclose this sensitive information in an anonymous Internet setting as compared to during a face-to-face interview [80,81]. This bias might be mitigated because interviewers were administered by highly trained MSM or MSM-friendly staff members. Moreover, participants in these studies reported high levels of another sensitive issue, condomless anal sex, which further suggests minimal bias due to social desirability (data not shown).

Another limitation is that the African studies were conducted with incentives for participation that were not included in AMIS, including in some cases HIV and biological testing, and therefore may have reached individuals who would have not otherwise participated. Participants in AMIS were highly educated and mostly of White, non-Hispanic race/ethnicity, suggesting limitations to generalizability to other MSM in the United States. It is also possible that AMIS participants, who were recruited from gay-related websites, were more actively involved in the gay community and more likely to be exposed to stigma. This is because participants in Sub-Saharan Africa were recruited using RDS (except in Swaziland), which has been shown to be capable of reaching MSM who are less engaged in the gay community and HIV testing [82]. Finally, the prevalence of missing values for AMIS-2015 (1.5%-10.7%

for each stigma item) were higher than those in Western and Southern Africa (0%-1%), likely because AMIS-2015 was a Web-based survey, whereas African data were collected during face-to-face interviews. However, this study is the first to our knowledge to make a direct comparison of sexual behavior stigma between MSM in the United States and Sub-Saharan Africa using consistent metrics.

## Conclusions

Sexual behavior stigma felt or experienced by MSM appears to be pervasive across the United States and Sub-Saharan Africa, and this has implications for both sexual and mental health. Given the growing desire to measure, quantify, and evaluate stigma toward key populations as part of the HIV response, this paper is timely in its comparison of prevalence of sexual behavior stigma across 3 widely different regions (the United States and Western and Southern Sub-Saharan Africa). It is important to note that results may be influenced by differences in sampling methodology across regions. However, the consistently observed high burden of stigma points to the need for immediate structural interventions to address each of the domains of sexual behavior stigma presented here; that is, stigma from family and friends, health care workers, and from broader society. These interventions will be critical for making a positive impact on the mental health of these men, and also for reducing the global sustained burden of HIV and other adverse health outcomes in this key population.

## Acknowledgments

AMIS-2015 was funded by a grant from the MAC AIDS Fund and by the National Institutes of Health (NIH) [P30AI050409] – the Emory Center for AIDS Research. Work in Togo and Burkina Faso was supported by Project SEARCH, which was funded by the US Agency for International Development under Contract GHH-I-00-07-00032-00 and by the President's Emergency Plan for AIDS Relief (PEPFAR). Work in Côte d'Ivoire was funded by the Global Fund to Fight AIDS, Tuberculosis and Malaria through the Government of Côte d'Ivoire National AIDS Control Program (PNPEC) contract to Enda Santé, an organization based in Senegal, and subcontracted for technical assistance to Johns Hopkins University. Work in Lesotho was funded by the US Agency for International Development (USAID, AID-674-A-00-00001), and implemented by Population Services International/Lesotho (PSI). Work in Nigeria was supported by a cooperative agreement (W81XWH-11-2-0174) between the Henry M. Jackson Foundation for the Advancement of Military Medicine, Inc., and the US Department of Defense (DOD). The Nigeria study is also supported by funds from the US NIH under Award No. R01MH099001-01, the US Military HIV Research Program (Grant No. W81XWH-07-2-0067), Fogarty AITRP (D43TW01041), and PEPFAR through cooperative agreement U2G IPS000651 from the United States Department of Health and Human Services (HHS)/Centers for Disease Control and Prevention (CDC), Global AIDS Program with the Institute of Human Virology Nigeria (IHVN). Work in Senegal was funded by USAID (AID-OAA-A-13-00089) and work in Swaziland was funded by PEPFAR through the USAID Swaziland (GHH-I-00-07-00032-00). Finally, this publication was made possible with help from the Johns Hopkins University Center for AIDS Research, an NIH funded program (P30AI094189), which is supported by the following NIH Cofunding and Participating Institutes and Centers: National Institute of Allergy and Infectious Diseases (NIAID), National Cancer Institute (NCI), National Institute of Child Health and Human Development (NICHD), National Heart, Lung, and Blood Institute (NHLBI), National Institute on Drug Abuse (NIDA), National Institute of Mental Health (NIMH), National Institute on Aging (NIA), Fogarty International Center (FIC), National Institute of General Medical Sciences (NIGMS), National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK), and the Office of AIDS Research (OAR). The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH. The funding sources had no role in the study design; in the collection, analysis and interpretation of data; in the writing of the articles; or in the decision to submit the manuscript for publication.

## Conflicts of Interest

None declared.

### Multimedia Appendix 1

Prevalence of sexual behavior stigma among MSM in AMIS-2015 by race/ethnicity.

[[PDF File \(Adobe PDF File\), 23KB - publichealth\\_v2i2e35\\_app1.pdf](#)]

---

### Multimedia Appendix 2

Prevalence of sexual behavior stigma among MSM in AMIS-2015 by United States region.

[[PDF File \(Adobe PDF File\), 23KB - publichealth\\_v2i2e35\\_app2.pdf](#)]

---

### Multimedia Appendix 3

Prevalence of sexual behavior stigma among MSM in AMIS-2015 by population density.

[[PDF File \(Adobe PDF File\), 22KB - publichealth\\_v2i2e35\\_app3.pdf](#)]

---

### Multimedia Appendix 4

Prevalence of sexual behavior stigma among MSM in AMIS-2015 by age group.

[[PDF File \(Adobe PDF File\), 22KB - publichealth\\_v2i2e35\\_app4.pdf](#)]

---

### Multimedia Appendix 5

Prevalence of sexual behavior stigma among MSM who disclosed same-sex behaviors to family vs. not disclosed, by United States/Africa region.

[[PDF File \(Adobe PDF File\), 21KB - publichealth\\_v2i2e35\\_app5.pdf](#)]

---

### Multimedia Appendix 6

Prevalence of sexual behavior stigma among MSM who disclosed same-sex behaviors to healthcare worker vs. not disclosed, by United States/Africa region.

[[PDF File \(Adobe PDF File\), 21KB - publichealth\\_v2i2e35\\_app6.pdf](#)]

---

### References

1. World Health Organization. Prevention and Treatment of HIV and Other Sexually Transmitted Infections Among Men Who Have Sex With Men and Transgender People: Recommendations for a Public Health Approach. 2011. URL: [http://www.who.int/hiv/pub/guidelines/msm\\_guidelines2011/en/](http://www.who.int/hiv/pub/guidelines/msm_guidelines2011/en/) [accessed 2016-07-11] [WebCite Cache ID 6ivSmgy4m]
2. Joint United Nations Programme on HIV/AIDS (UNAIDS). Action on Social Drivers to End AIDS and Extreme Poverty: An Expert Consultation on HIV/AIDS and the Global Health and Development Agendas. Washington, D.C: World Bank Main Complex; 2014 Jan 10. URL: [https://www.health.ny.gov/diseases/aids/ending\\_the\\_epidemic/docs/key\\_resources/housing\\_and\\_supportive\\_services/social\\_drivers\\_to\\_end\\_aids.pdf](https://www.health.ny.gov/diseases/aids/ending_the_epidemic/docs/key_resources/housing_and_supportive_services/social_drivers_to_end_aids.pdf) [accessed 2016-07-11] [WebCite Cache ID 6ivT9FBxP]
3. White House Office of National AIDS Policy. National HIV/AIDS Strategy for the United States: Updated to 2020. 2015. URL: <https://www.aids.gov/federal-resources/national-hiv-aids-strategy/overview/> [accessed 2016-07-11] [WebCite Cache ID 6ivTNgZjT]
4. Beyrer C, Baral SD, van GF, Goodreau SM, Chariyalertsak S, Wirtz AL, et al. Global epidemiology of HIV infection in men who have sex with men. *Lancet* 2012;380:367-377 [FREE Full text] [doi: [10.1016/S0140-6736\(12\)60821-6](https://doi.org/10.1016/S0140-6736(12)60821-6)] [Medline: [22819660](https://pubmed.ncbi.nlm.nih.gov/22819660/)]
5. Beyrer C, Sullivan P, Sanchez J, Baral SD, Collins C, Wirtz AL, et al. The increase in global HIV epidemics in MSM. *AIDS* 2013;27:2665-2678. [doi: [10.1097/01.aids.0000432449.30239.fe](https://doi.org/10.1097/01.aids.0000432449.30239.fe)] [Medline: [23842129](https://pubmed.ncbi.nlm.nih.gov/23842129/)]
6. Hess K, Hu X, Lansky A, Mermin J, Hall H. Estimating the lifetime risk of a diagnosis of HIV infection in the United States. 2016 Presented at: Conference on Retroviruses and Opportunistic Infections (CROI); February 23, 2016; Boston, MA URL: <http://www.croiconference.org/sessions/estimating-lifetime-risk-diagnosis-hiv-infection-united-states>
7. Baral S, Sifakis F, Cleghorn F, Beyrer C. Elevated risk for HIV infection among men who have sex with men in low- and middle-income countries 2000-2006: a systematic review. *PLoS Med* 2007;4:e339 [FREE Full text] [doi: [10.1371/journal.pmed.0040339](https://doi.org/10.1371/journal.pmed.0040339)] [Medline: [18052602](https://pubmed.ncbi.nlm.nih.gov/18052602/)]
8. Baral S. HIV risks and vulnerabilities among men who have sex with men across Sub-Saharan Africa. 2015 Presented at: Conference on Retroviruses and Opportunistic Infections (CROI); 2015; Seattle, Washington.
9. van Griensven F, de Lind van Wijngaarden JW, Baral S, Grulich A. The global epidemic of HIV infection among men who have sex with men. *Curr Opin HIV AIDS* 2009;4:300-307. [doi: [10.1097/COH.0b013e32832c3bb3](https://doi.org/10.1097/COH.0b013e32832c3bb3)] [Medline: [19532068](https://pubmed.ncbi.nlm.nih.gov/19532068/)]



10. Brown L, Macintyre K, Trujillo L. Interventions to reduce HIV/AIDS stigma: what have we learned? *AIDS Educ Prev* 2003;15:49-69. [Medline: [12627743](#)]
11. Knox J, Sandfort T, Yi H, Reddy V, Maimane S. Social vulnerability and HIV testing among South African men who have sex with men. *Int J STD AIDS* 2011;22:709-713 [FREE Full text] [doi: [10.1258/ijsa.2011.010350](#)] [Medline: [22174050](#)]
12. Hladik W, Barker J, Ssenkusu JM, Opio A, Tappero JW, Hakim A, Crane Survey Group. HIV infection among men who have sex with men in Kampala, Uganda--a respondent driven sampling survey. *PLoS One* 2012;7:e38143 [FREE Full text] [doi: [10.1371/journal.pone.0038143](#)] [Medline: [22693590](#)]
13. Hu Y, Lu H, Raymond HF, Sun Y, Sun J, Jia Y, et al. Measures of condom and safer sex social norms and stigma towards HIV/AIDS among Beijing MSM. *AIDS Behav* 2014;18:1068-1074 [FREE Full text] [doi: [10.1007/s10461-013-0609-7](#)] [Medline: [24057931](#)]
14. Choi K, Hudes ES, Steward WT. Social discrimination, concurrent sexual partnerships, and HIV risk among men who have sex with men in Shanghai, China. *AIDS Behav* 2008;12(Suppl):S71-S77. [doi: [10.1007/s10461-008-9394-0](#)] [Medline: [18427972](#)]
15. Cochran SD, Mays VM, Sullivan JG. Prevalence of mental disorders, psychological distress, and mental health services use among lesbian, gay, and bisexual adults in the United States. *J Consult Clin Psychol* 2003;71:53-61 [FREE Full text] [Medline: [12602425](#)]
16. Stahlman S, Grosso A, Ketende S, Sweitzer S, Mothopeng T, Taruberekera N, et al. Depression and social stigma among MSM in Lesotho: implications for HIV and sexually transmitted infection prevention. *AIDS Behav* 2015;19:1460-1469. [doi: [10.1007/s10461-015-1094-y](#)] [Medline: [25969182](#)]
17. Stoloff K, Joska JA, Feast D, De SG, Hugo J, Struthers H, et al. A description of common mental disorders in men who have sex with men (MSM) referred for assessment and intervention at an MSM clinic in Cape Town, South Africa. *AIDS Behav* 2013;Suppl 1:S77-S81. [doi: [10.1007/s10461-013-0430-3](#)] [Medline: [23532397](#)]
18. Baral SD, Ketende S, Mnisi Z, Mabuza X, Grosso A, Sithole B, et al. A cross-sectional assessment of the burden of HIV and associated individual- and structural-level characteristics among men who have sex with men in Swaziland. *J Int AIDS Soc* 2013;Suppl 3:18768 [FREE Full text] [Medline: [24321117](#)]
19. The Foundation for AIDS Research (amfAR), Johns Hopkins Bloomberg School of Public Health. Achieving an AIDS-Free Generation for Gay Men and Other MSM: Financing Implementation of HIV Programs Targeting MSM. 2012. URL: [http://www.amfar.org/uploadedFiles/amfar.org/In\\_The\\_Community/Publications/MSM-GlobalRept2012.pdf](http://www.amfar.org/uploadedFiles/amfar.org/In_The_Community/Publications/MSM-GlobalRept2012.pdf) [accessed 2016-07-11] [WebCite Cache ID 6ivXfnfkl]
20. Pachankis JE, Hatzenbuehler ML, Hickson F, Weatherburn P, Berg RC, Marcus U, et al. Hidden from health: structural stigma, sexual orientation concealment, and HIV across 38 countries in the European MSM Internet Survey. *AIDS* 2015;29:1239-1246. [doi: [10.1097/QAD.0000000000000724](#)] [Medline: [26035323](#)]
21. Ayala G, Makofane K, Santos G, Beck J, Do TD, Hebert P, et al. Access to basic HIV-related services and PrEP acceptability among men who have sex with men worldwide: barriers, facilitators, and implications for combination prevention. *J Sex Transm Dis* 2013;2013:953123 [FREE Full text] [doi: [10.1155/2013/953123](#)] [Medline: [26316968](#)]
22. Baral S, Holland CE, Shannon K, Logie C, Semugoma P, Sithole B, et al. Enhancing benefits or increasing harms: community responses for HIV among men who have sex with men, transgender women, female sex workers, and people who inject drugs. *J Acquir Immune Defic Syndr* 2014;Suppl 3:S319-S328. [doi: [10.1097/QAI.0000000000000233](#)] [Medline: [25007203](#)]
23. Schwartz S, Nowak R, Orazulike I, Keshinro B, Ake J, Kennedy S, et al. The immediate effect of the Same-Sex Marriage Prohibition Act on stigma, discrimination, and engagement on HIV prevention and treatment services in men who have sex with men in Nigeria: analysis of prospective data from the TRUST cohort. *Lancet HIV* 2015;2:e299-e306 [FREE Full text] [doi: [10.1016/S2352-3018\(15\)00078-8](#)] [Medline: [26125047](#)]
24. Oldenburg CE, Perez-Brumer AG, Hatzenbuehler ML, Krakower D, Novak DS, Mimiaga MJ, et al. State-level structural sexual stigma and HIV prevention in a national online sample of HIV-uninfected MSM in the United States. *AIDS* 2015;29:837-845 [FREE Full text] [doi: [10.1097/QAD.0000000000000622](#)] [Medline: [25730508](#)]
25. Hatzenbuehler ML, McLaughlin KA. Structural stigma and hypothalamic-pituitary-adrenocortical axis reactivity in lesbian, gay, and bisexual young adults. *Ann Behav Med* 2014;47:39-47 [FREE Full text] [doi: [10.1007/s12160-013-9556-9](#)] [Medline: [24154988](#)]
26. Centers for Disease Control and Prevention (CDC). HIV infections attributed to male-to-male sexual contact - metropolitan statistical areas, United States and Puerto Rico, 2010. *MMWR Morb Mortal Wkly Rep* 2012;61:962-966 [FREE Full text] [Medline: [23190569](#)]
27. Prejean J, Tang T, Hall HI. HIV diagnoses and prevalence in the southern region of the United States, 2007-2010. *J Community Health* 2013;38:414-426. [doi: [10.1007/s10900-012-9633-1](#)] [Medline: [23179388](#)]
28. Stall R, Duran L, Wisniewski SR, Friedman MS, Marshal MP, McFarland W, et al. Running in place: implications of HIV incidence estimates among urban men who have sex with men in the United States and other industrialized countries. *AIDS Behav* 2009;13:615-629 [FREE Full text] [doi: [10.1007/s10461-008-9509-7](#)] [Medline: [19205867](#)]
29. Millett GA, Peterson JL, Wolitski RJ, Stall R. Greater risk for HIV infection of black men who have sex with men: a critical literature review. *Am J Public Health* 2006;96:1007-1019. [doi: [10.2105/AJPH.2005.066720](#)] [Medline: [16670223](#)]

30. Millett GA, Flores SA, Peterson JL, Bakeman R. Explaining disparities in HIV infection among black and white men who have sex with men: a meta-analysis of HIV risk behaviors. *AIDS* 2007;21:2083-2091. [doi: [10.1097/QAD.0b013e3282e9a64b](https://doi.org/10.1097/QAD.0b013e3282e9a64b)] [Medline: [17885299](https://pubmed.ncbi.nlm.nih.gov/17885299/)]
31. Centers for Disease Control and Prevention (CDC). HIV Incidence. 2010. URL: <http://www.cdc.gov/hiv/statistics/surveillance/incidence.html> [accessed 2016-01-20] [WebCite Cache ID 6ivYI0I0z]
32. Matthews DD, Herrick AL, Coulter RWS, Friedman MR, Mills TC, Eaton LA, et al. Running backwards: consequences of current HIV incidence rates for the next generation of Black MSM in the United States. *AIDS Behav* 2016;20:7-16. [doi: [10.1007/s10461-015-1158-z](https://doi.org/10.1007/s10461-015-1158-z)] [Medline: [26267251](https://pubmed.ncbi.nlm.nih.gov/26267251/)]
33. Millett GA, Peterson JL, Flores SA, Hart TA, Jeffries WL, Wilson PA, et al. Comparisons of disparities and risks of HIV infection in black and other men who have sex with men in Canada, UK, and USA: a meta-analysis. *Lancet* 2012;380:341-348. [doi: [10.1016/S0140-6736\(12\)60899-X](https://doi.org/10.1016/S0140-6736(12)60899-X)] [Medline: [22819656](https://pubmed.ncbi.nlm.nih.gov/22819656/)]
34. Poteat T, Reischer SL, Radix A. HIV epidemics among transgender women. *Curr Opin HIV AIDS* 2014;9:168-173. [doi: [10.1097/COH.000000000000030](https://doi.org/10.1097/COH.000000000000030)] [Medline: [24322537](https://pubmed.ncbi.nlm.nih.gov/24322537/)]
35. De Santis JP. HIV infection risk factors among male-to-female transgender persons: a review of the literature. *J Assoc Nurses AIDS Care* 2009;20:362-372. [doi: [10.1016/j.jana.2009.06.005](https://doi.org/10.1016/j.jana.2009.06.005)] [Medline: [19732695](https://pubmed.ncbi.nlm.nih.gov/19732695/)]
36. Balaji A, Smith J, Bowles K, Paz-Bailey G, NHBS study group. Association between enacted stigma and HIV related risk behavior among MSM, National HIV Behavioral Surveillance System. 2011 Presented at: Conference on Retroviruses and Opportunistic Infections (CROI); February 23-26, 2015; Seattle, Washington URL: <http://www.croiconference.org/sites/default/files/posters-2015/1058.pdf>
37. Cange CW, LeBreton M, Billong S, Saylor K, Tamoufe U, Papworth E, et al. Influence of stigma and homophobia on mental health and on the uptake of HIV/sexually transmissible infection services for Cameroonian men who have sex with men. *Sex Health* 2015. [doi: [10.1071/SH15001](https://doi.org/10.1071/SH15001)] [Medline: [26117222](https://pubmed.ncbi.nlm.nih.gov/26117222/)]
38. Fay H, Baral SD, Trapence G, Motimedi F, Umar E, Ipinge S, et al. Stigma, health care access, and HIV knowledge among men who have sex with men in Malawi, Namibia, and Botswana. *AIDS Behav* 2011;15:1088-1097. [doi: [10.1007/s10461-010-9861-2](https://doi.org/10.1007/s10461-010-9861-2)] [Medline: [21153432](https://pubmed.ncbi.nlm.nih.gov/21153432/)]
39. Risher K, Adams D, Sithole B, Ketende S, Kennedy C, Mnisi Z, et al. Sexual stigma and discrimination as barriers to seeking appropriate healthcare among men who have sex with men in Swaziland. *J Int AIDS Soc* 2013;16(Suppl 2):18715 [FREE Full text] [Medline: [24242263](https://pubmed.ncbi.nlm.nih.gov/24242263/)]
40. Fitzgerald-Husek A, Van Wert M, Rosman L, Bowes A, Grosso A, Baral S. Measuring stigma related to sex worker identity and MSM identity: a systematic review. 2014 Nov 18 Presented at: APHA 142nd Annual Meeting and Expo; November 15-19, 2014; New Orleans, LA URL: <https://apha.confex.com/apha/142am/webprogram/Paper312252.html>
41. Sanchez T, Sineath R, Kahle E, Tregear S, Sullivan P. The annual american men's internet survey of behaviors of men who have sex with men in the United States: protocol and key indicators report 2013. *JMIR Public Health Surveill* 2015;1:e3 [FREE Full text] [doi: [10.2196/publichealth.4314](https://doi.org/10.2196/publichealth.4314)] [Medline: [27227126](https://pubmed.ncbi.nlm.nih.gov/27227126/)]
42. Emory University. American Men's Internet Survey (AMIS): Online HIV Behavioral Survey of Men Who Have Sex With Men. Atlanta, GA: PRISM Health, Emory University; 2015. URL: [http://www.cfar.emory.edu/downloads/news/2016/AMIS-2014\\_National\\_Report\\_&\\_Survey.pdf](http://www.cfar.emory.edu/downloads/news/2016/AMIS-2014_National_Report_&_Survey.pdf) [accessed 2016-07-11]
43. Stahlman S, Bechtold K, Sweitzer S, Mothopeng T, Tarubekera N, Nkonyana J, et al. Sexual identity stigma and social support among men who have sex with men in Lesotho: a qualitative analysis. *Reprod Health Matters* 2015;23:127-135. [doi: [10.1016/j.rhm.2015.11.017](https://doi.org/10.1016/j.rhm.2015.11.017)] [Medline: [26719004](https://pubmed.ncbi.nlm.nih.gov/26719004/)]
44. Stahlman S, Grosso A, Ketende S, Mothopeng T, Tarubekera N, Nkonyana J, et al. Characteristics of men who have sex with men in southern Africa who seek sex online: a cross-sectional study. *J Med Internet Res* 2015;17:e129 [FREE Full text] [doi: [10.2196/jmir.4230](https://doi.org/10.2196/jmir.4230)] [Medline: [26006788](https://pubmed.ncbi.nlm.nih.gov/26006788/)]
45. Stahlman S, Holland C, Ketende S, Lith L, Kochelani D, Mnisi Z, et al. Stigma and openness about sexual identity among men who have sex with men: a latent class analysis. 2016 Presented at: Conference on Retroviruses and Opportunistic Infections (CROI); February 22-25, 2016; Boston, MA URL: <http://www.croiconference.org/sessions/stigma-and-openness-about-sexual-identity-among-msm-latent-class-analysis>
46. Heckathorn D. Respondent-driven sampling: a new approach to the study of hidden populations. *Social Problems* 1997;44:174-199. [doi: [10.2307/3096941](https://doi.org/10.2307/3096941)]
47. Biernacki P, Waldorf D. Snowball sampling: problems and techniques of chain referral sampling. *Sociological Methods & Research* 1981;10:141-163. [doi: [10.1177/004912418101000205](https://doi.org/10.1177/004912418101000205)]
48. International Lesbian Gay Bisexual Trans and Intersex Association. State Sponsored Homophobia Report. 2015. URL: <http://ilga.org/what-we-do/state-sponsored-homophobia-report/> [accessed 2016-07-12] [WebCite Cache ID 6iwgGWsjj]
49. Itaborahy L, Zhu J. A World Survey of Laws: Criminalisation, Protection and Recognition of Same-Sex Love.: International Lesbian Gay Bisexual Trans and Intersex Association; 2013. State-Sponsored Homophobia URL: [http://old.ilga.org/Statehomophobia/ILGA\\_State\\_Sponsored\\_Homophobia\\_2013.pdf](http://old.ilga.org/Statehomophobia/ILGA_State_Sponsored_Homophobia_2013.pdf) [accessed 2016-07-12] [WebCite Cache ID 6iwgRTnTs]
50. Itaborahy LP, Zhu J. A World Survey of Laws: Criminalisation, Protection and Recognition of Same-Sex Love.: International Lesbian Gay Bisexual Trans and Intersex Association; 2014. State-Sponsored Homophobia URL: [http://old.ilga.org/Statehomophobia/ILGA\\_SSHR\\_2014\\_Eng.pdf](http://old.ilga.org/Statehomophobia/ILGA_SSHR_2014_Eng.pdf) [accessed 2016-07-12] [WebCite Cache ID 6iwgEQZ6]

51. Yebei VN, Fortenberry JD, Ayuku DO. Felt stigma among people living with HIV/AIDS in rural and urban Kenya. *Afr Health Sci* 2008;8:97-102 [FREE Full text] [Medline: [19357758](#)]
52. Halkitis PN, Perez-Figueroa RE, Carreiro T, Kingdon MJ, Kupprat SA, Eddy J. Psychosocial burdens negatively impact HIV antiretroviral adherence in gay, bisexual, and other men who have sex with men aged 50 and older. *AIDS Care* 2014;26:1426-1434 [FREE Full text] [doi: [10.1080/09540121.2014.921276](#)] [Medline: [24865599](#)]
53. Lépine A, Terris-Prestholt F, Vickerman P. Determinants of HIV testing among Nigerian couples: a multilevel modelling approach. *Health Policy and Planning* 2014;30:579-592. [doi: [10.1093/heapol/czu036](#)]
54. Kahana SY, Jenkins RA, Bruce D, Fernandez MI, Hightow-Weidman LB, Bauermeister JA, Adolescent Medicine Trials Network for HIV/AIDS Interventions. Structural determinants of antiretroviral therapy use, HIV care attendance, and viral suppression among adolescents and young adults living with HIV. *PLoS One* 2016;11:e0151106 [FREE Full text] [doi: [10.1371/journal.pone.0151106](#)] [Medline: [27035905](#)]
55. Wei C, Cheung DH, Yan H, Li J, Shi L, Raymond HF. The impact of homophobia and HIV stigma on HIV testing uptake among chinese men who have sex with men: a mediation analysis. *J Acquir Immune Defic Syndr* 2016;71:87-93. [doi: [10.1097/QAI.0000000000000815](#)] [Medline: [26334742](#)]
56. Secor AM, Wahome E, Micheni M, Rao D, Simoni JM, Sanders EJ, et al. Depression, substance abuse and stigma among men who have sex with men in coastal Kenya. *AIDS* 2015;Suppl 3:S251-S259 [FREE Full text] [doi: [10.1097/QAD.0000000000000846](#)] [Medline: [26562814](#)]
57. Choi K, Steward WT, Miège P, Hudes E, Gregorich SE. Sexual stigma, coping styles, and psychological distress: a longitudinal study of men who have sex with men in Beijing, China. *Arch Sex Behav* 2015. [doi: [10.1007/s10508-015-0640-z](#)] [Medline: [26679303](#)]
58. Logie CH, Newman PA, Chakrapani V, Shunmugam M. Adapting the minority stress model: associations between gender non-conformity stigma, HIV-related stigma and depression among men who have sex with men in South India. *Soc Sci Med* 2012;74:1261-1268. [doi: [10.1016/j.socscimed.2012.01.008](#)] [Medline: [22401646](#)]
59. Zervoulis K, Lyons E, Dinos S. Stigma and self-esteem across societies: avoiding blanket psychological responses to gay men experiencing homophobia. *BJPsych Bull* 2015;39:167-173 [FREE Full text] [doi: [10.1192/pb.bp.114.048421](#)] [Medline: [26755948](#)]
60. Yannessa JF, Reece M, Basta TB. HIV provider perspectives: the impact of stigma on substance abusers living with HIV in a rural area of the United States. *AIDS Patient Care STDS* 2008;22:669-675. [doi: [10.1089/apc.2007.0151](#)] [Medline: [18627281](#)]
61. Rost K, Smith GR, Taylor JL. Rural-urban differences in stigma and the use of care for depressive disorders. *J Rural Health* 1993;9:57-62. [Medline: [10124199](#)]
62. Glick SN, Golden MR. Persistence of racial differences in attitudes toward homosexuality in the United States. *J Acquir Immune Defic Syndr* 2010;55:516-523 [FREE Full text] [doi: [10.1097/QAI.0b013e3181f275e0](#)] [Medline: [20838226](#)]
63. Lewis GB. Black-White differences in attitudes toward homosexuality and gay rights. *Public Opinion Quarterly* 2003;67:59-78. [doi: [10.1086/346009](#)]
64. Quinn K, Dickson-Gomez J, DiFranceisco W, Kelly JA, St Lawrence JS, Amirkhanian YA, et al. Correlates of internalized homonegativity among black men who have sex with men. *AIDS Educ Prev* 2015;27:212-226 [FREE Full text] [doi: [10.1521/aeap.2015.27.3.212](#)] [Medline: [26010313](#)]
65. Garofalo R, Hotton AL, Kuhns LM, Gratzner B, Mustanski B. Incidence of HIV infection and sexually transmitted infections and related risk factors among very young men who have sex with men. *J Acquir Immune Defic Syndr* 2016. [doi: [10.1097/QAI.0000000000000933](#)] [Medline: [26745827](#)]
66. Bedoya CA, Mimiaga MJ, Beauchamp G, Donnell D, Mayer KH, Safren SA. Predictors of HIV transmission risk behavior and seroconversion among Latino men who have sex with men in Project EXPLORE. *AIDS Behav* 2012;16:608-617 [FREE Full text] [doi: [10.1007/s10461-011-9911-4](#)] [Medline: [21390540](#)]
67. Mustanski B, Newcomb M, Du Bois SN, Garcia S, Grov C. HIV in young men who have sex with men: a review of epidemiology, risk and protective factors, and interventions. *J Sex Res* 2011;48:218-253 [FREE Full text] [doi: [10.1080/00224499.2011.558645](#)] [Medline: [21409715](#)]
68. Baral SD, Grosso A, Holland C, Papworth E. The epidemiology of HIV among men who have sex with men in countries with generalized HIV epidemics. *Curr Opin HIV AIDS* 2014;9:156-167. [doi: [10.1097/COH.0000000000000037](#)] [Medline: [24445371](#)]
69. Sullivan PS, Rosenberg ES, Sanchez TH, Kelley CF, Luisi N, Cooper HL, et al. Explaining racial disparities in HIV incidence in black and white men who have sex with men in Atlanta, GA: a prospective observational cohort study. *Ann Epidemiol* 2015;25:445-454. [doi: [10.1016/j.annepidem.2015.03.006](#)] [Medline: [25911980](#)]
70. McAdams-Mahmoud A, Stephenson R, Rentsch C, Cooper H, Arriola KJ, Jobson G, et al. Minority stress in the lives of men who have sex with men in Cape Town, South Africa. *J Homosex* 2014;61:847-867. [doi: [10.1080/00918369.2014.870454](#)] [Medline: [24392722](#)]
71. Meyer IH. Prejudice, social stress, and mental health in lesbian, gay, and bisexual populations: conceptual issues and research evidence. *Psychol Bull* 2003;129:674-697 [FREE Full text] [doi: [10.1037/0033-2909.129.5.674](#)] [Medline: [12956539](#)]

72. D'augelli A, Grossman A. Disclosure of sexual orientation, victimization, and mental health among lesbian, gay, and bisexual older adults. *Journal of Interpersonal Violence* 2001;16:1008-1027. [doi: [10.1177/088626001016010003](https://doi.org/10.1177/088626001016010003)]
73. Cahill S, Valadéz R, Ibarrola S. Community-based HIV prevention interventions that combat anti-gay stigma for men who have sex with men and for transgender women. *J Public Health Policy* 2013;34:69-81. [doi: [10.1057/jphp.2012.59](https://doi.org/10.1057/jphp.2012.59)] [Medline: [23151921](https://pubmed.ncbi.nlm.nih.gov/23151921/)]
74. Charurat ME, Emmanuel B, Akolo C, Keshinro B, Nowak RG, Kennedy S, et al. Uptake of treatment as prevention for HIV and continuum of care among HIV-positive men who have sex with men in Nigeria. *J Acquir Immune Defic Syndr* 2015;Suppl 2:S114-S123 [FREE Full text] [doi: [10.1097/QAI.0000000000000439](https://doi.org/10.1097/QAI.0000000000000439)] [Medline: [25723975](https://pubmed.ncbi.nlm.nih.gov/25723975/)]
75. McNaghten A, Kearns R, Siegler AJ, Phaswana-Mafuya N, Bekker L, Stephenson R, et al. Methods for prevention packages program project protocol: pilot study of hiv prevention interventions for men who have sex with men in South Africa. *JMIR Res Protoc* 2014;3:e55 [FREE Full text] [doi: [10.2196/resprot.3737](https://doi.org/10.2196/resprot.3737)] [Medline: [25325296](https://pubmed.ncbi.nlm.nih.gov/25325296/)]
76. van der Elst EM, Kombo B, Gichuru E, Omar A, Musyoki H, Graham SM, et al. The green shoots of a novel training programme: progress and identified key actions to providing services to MSM at Kenyan health facilities. *J Int AIDS Soc* 2015;18:20226 [FREE Full text] [Medline: [26493863](https://pubmed.ncbi.nlm.nih.gov/26493863/)]
77. Colchero MA, Bautista-Arredondo S, Cortés-Ortiz MA, Romero-Martinez M, Salas J, Sosa-Rubí SG, et al. Impact and economic evaluations of a combination prevention programme for men who have sex with men in Mexico. *AIDS* 2016;30:293-300. [doi: [10.1097/QAD.0000000000000925](https://doi.org/10.1097/QAD.0000000000000925)] [Medline: [26691549](https://pubmed.ncbi.nlm.nih.gov/26691549/)]
78. Henderson M. *On the Ground: Programmes Serving the Needs of Key Populations.*: World Health Organization; 2014. URL: [http://apps.who.int/iris/bitstream/10665/145002/1/WHO\\_HIV\\_2014.50\\_eng.pdf](http://apps.who.int/iris/bitstream/10665/145002/1/WHO_HIV_2014.50_eng.pdf) [accessed 2016-07-12] [WebCite Cache ID 6iwhyDjzo]
79. Toomey R, Ryan C, Diaz R, Russell S. High school gay-straight alliances (GSAs) and young adult well-being: an examination of GSA presence, participation, and perceived effectiveness. *Applied Developmental Science* 2011;15:175-185. [doi: [10.1080/10888691.2011.607378](https://doi.org/10.1080/10888691.2011.607378)]
80. Adebajo S, Obianwu O, Eluwa G, Vu L, Oginni A, Tun W, et al. Comparison of audio computer assisted self-interview and face-to-face interview methods in eliciting HIV-related risks among men who have sex with men and men who inject drugs in Nigeria. *PLoS One* 2014;9:e81981 [FREE Full text] [doi: [10.1371/journal.pone.0081981](https://doi.org/10.1371/journal.pone.0081981)] [Medline: [24416134](https://pubmed.ncbi.nlm.nih.gov/24416134/)]
81. Castro R. Inconsistent respondents and sensitive questions. *Field Methods* 2012;25:283-298. [doi: [10.1177/1525822x12466988](https://doi.org/10.1177/1525822x12466988)]
82. Stahlman S, Johnston LG, Yah C, Ketende S, Maziya S, Trapence G, et al. Respondent-driven sampling as a recruitment method for men who have sex with men in southern sub-Saharan Africa: a cross-sectional analysis by wave. *Sex Transm Infect* 2016;92:292-298. [doi: [10.1136/sextrans-2015-052184](https://doi.org/10.1136/sextrans-2015-052184)] [Medline: [26424712](https://pubmed.ncbi.nlm.nih.gov/26424712/)]

## Abbreviations

**CBOs:** community-based organizations

**CI:** confidence interval

**HIV:** human immunodeficiency virus

**IQR:** interquartile range

**IRB:** institutional review board

**MSM:** men who have sex with men

**PR:** prevalence ratio

**RDS:** respondent driven sampling

*Edited by G Eysenbach; submitted 31.03.16; peer-reviewed by S LeGrand, C Oldenburg; comments to author 03.05.16; revised version received 20.05.16; accepted 10.06.16; published 26.07.16.*

*Please cite as:*

Stahlman S, Sanchez TH, Sullivan PS, Ketende S, Lyons C, Charurat ME, Drame FM, Diouf D, Ezouatchi R, Kouanda S, Anato S, Mothopeng T, Mnisi Z, Baral SD

*The Prevalence of Sexual Behavior Stigma Affecting Gay Men and Other Men Who Have Sex with Men Across Sub-Saharan Africa and in the United States*

*JMIR Public Health Surveill* 2016;2(2):e35

URL: <http://publichealth.jmir.org/2016/2/e35/>

doi: [10.2196/publichealth.5824](https://doi.org/10.2196/publichealth.5824)

PMID: [27460627](https://pubmed.ncbi.nlm.nih.gov/27460627/)



©Shauna Stahlman, Travis Howard Sanchez, Patrick Sean Sullivan, Sosthenes Ketende, Carrie Lyons, Manhattan E Charurat, Fatou Maria Drame, Daouda Diouf, Rebecca Ezouatchi, Seni Kouanda, Simplicie Anato, Tampose Mothopeng, Zandile Mnisi, Stefan David Baral. Originally published in JMIR Public Health and Surveillance (<http://publichealth.jmir.org>), 26.07.2016. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Public Health and Surveillance, is properly cited. The complete bibliographic information, a link to the original publication on <http://publichealth.jmir.org>, as well as this copyright and license information must be included.



## Short Paper

# High Burden of Unrecognized Atrial Fibrillation in Rural India: An Innovative Community-Based Cross-Sectional Screening Program

Apurv Soni<sup>1</sup>, BA; Allison Earon<sup>1</sup>, MPH; Anna Handorf<sup>1</sup>, BA; Nisha Fahey<sup>1</sup>, BA; Kandarp Talati<sup>2</sup>, MBA; John Bostrom<sup>1</sup>, BA; Ki Chon<sup>3</sup>, PhD; Craig Napolitano<sup>1</sup>, MD; Michael Chin<sup>1</sup>, MD; John Sullivan<sup>1</sup>, BA; Shyamsundar Raithatha<sup>2</sup>, MD; Robert Goldberg<sup>1</sup>, PhD; Somashekhar Nimbalkar<sup>2</sup>, MD; Jeroan Allison<sup>1</sup>, MScEpi, MD; Sunil Thanvi<sup>2</sup>, DM, MD; David McManus<sup>1</sup>, MScI, MD

<sup>1</sup>University of Massachusetts Medical School, Worcester, MA, United States

<sup>2</sup>Pramukhswami Medical College, Karamsad, India

<sup>3</sup>University of Connecticut, Storrs, CT, United States

**Corresponding Author:**

Apurv Soni, BA

University of Massachusetts Medical School

55 Lake Ave North

Worcester, MA,

United States

Phone: 1 508 856 4265

Fax: 1 508 856 4267

Email: [apurv.soni@umassmed.edu](mailto:apurv.soni@umassmed.edu)

## Abstract

**Background:** Atrial fibrillation, the world's most common arrhythmia, is a leading risk factor for stroke, a disease striking nearly 1.6 million Indians annually. Early detection and management of atrial fibrillation is a promising opportunity to prevent stroke but widespread screening programs in limited resource settings using conventional methods is difficult and costly.

**Objective:** The objective of this study is to screen people for atrial fibrillation in rural western India using a US Food and Drug Administration-approved single-lead electrocardiography device, Alivecor.

**Methods:** Residents from 6 villages in Anand District, Gujarat, India, comprised the base population. After obtaining informed consent, a team of trained research coordinators and community health workers enrolled a total of 354 participants aged 50 years and older and screened them at their residences using Alivecor for 2 minutes on 5 consecutive days over a period of 6 weeks beginning June, 2015.

**Results:** Almost two-thirds of study participants were 55 years or older, nearly half were female, one-third did not receive any formal education, and more than one-half were from households earning less than US \$2 per day. Twelve participants screened positive for atrial fibrillation yielding a sample prevalence of 5.1% (95% CI 2.7-8.7). Only one participant had persistent atrial fibrillation throughout all of the screenings, and 9 screened positive only once.

**Conclusions:** Our study suggests a prevalence of atrial fibrillation in this Indian region (5.1%) that is markedly higher than has been previously reported in India and similar to the prevalence estimates reported in studies of persons from North America and Europe. Historically low reported burden of atrial fibrillation among individuals from low and middle-income countries may be due to a lack of routine screening. Mobile technologies may help overcome resource limitations for atrial fibrillation screening in underserved and low-resource settings.

(*JMIR Public Health Surveill* 2016;2(2):e159) doi:[10.2196/publichealth.6517](https://doi.org/10.2196/publichealth.6517)

**KEYWORDS**

atrial fibrillation; India; screening; mobile technology; community health workers

## Introduction

Atrial fibrillation is the world's most common cardiac arrhythmia and, if untreated, increases the risk of stroke by upwards of five-fold [1]. Atrial fibrillation-related complications, particularly stroke, have reached epidemic proportions in low and middle-income countries. This is particularly true in India, where approximately 1.6 million persons suffer a stroke annually [2]. A growing number of people in India are affected by risk factors for atrial fibrillation, including hypertension and diabetes mellitus [3], and the contribution of atrial fibrillation to the ongoing stroke epidemic in India is unclear and understudied [4]. In India, where the majority of health care costs are out of pocket [5], routine evaluations using conventional electrocardiography (ECG) to diagnose atrial fibrillation are not standard of care. Therefore, an understanding of the atrial fibrillation epidemiology becomes dependent on systematic screening programs. Single-time, point-of-care screening programs face difficulties of their own because of the paroxysmal and minimally symptomatic nature of the majority of atrial fibrillation cases.

Here we report findings of a study to screen people for atrial fibrillation in rural western India using a US Food and Drug Administration (FDA)-approved single-lead ECG device, Alivecor, to overcome traditional constraints of dysrhythmia screening [6].

## Methods

Residents from 6 different villages in Anand District, Gujarat, India, comprised the base population. These 6 villages were randomly selected from a list of 30 villages where our community health workers are present. Trained research coordinators worked with the community health workers who were familiar with the layout of their respective villages and enrolled 60 participants from each village. Villages in India are typically organized by occupation-based colonies (fariyahs), and an equal number of participants were recruited from all fariyahs. The residents of every third house in each fariyah were approached for enrollment through the use of a systematic random sample. After obtaining informed consent, a team of trained research coordinators and community health workers enrolled a total of 355 participants aged 50 years and older to participate in the study.

The study included two components: (1) screening using FDA-approved single-lead ECG device, Alivecor, and (2)

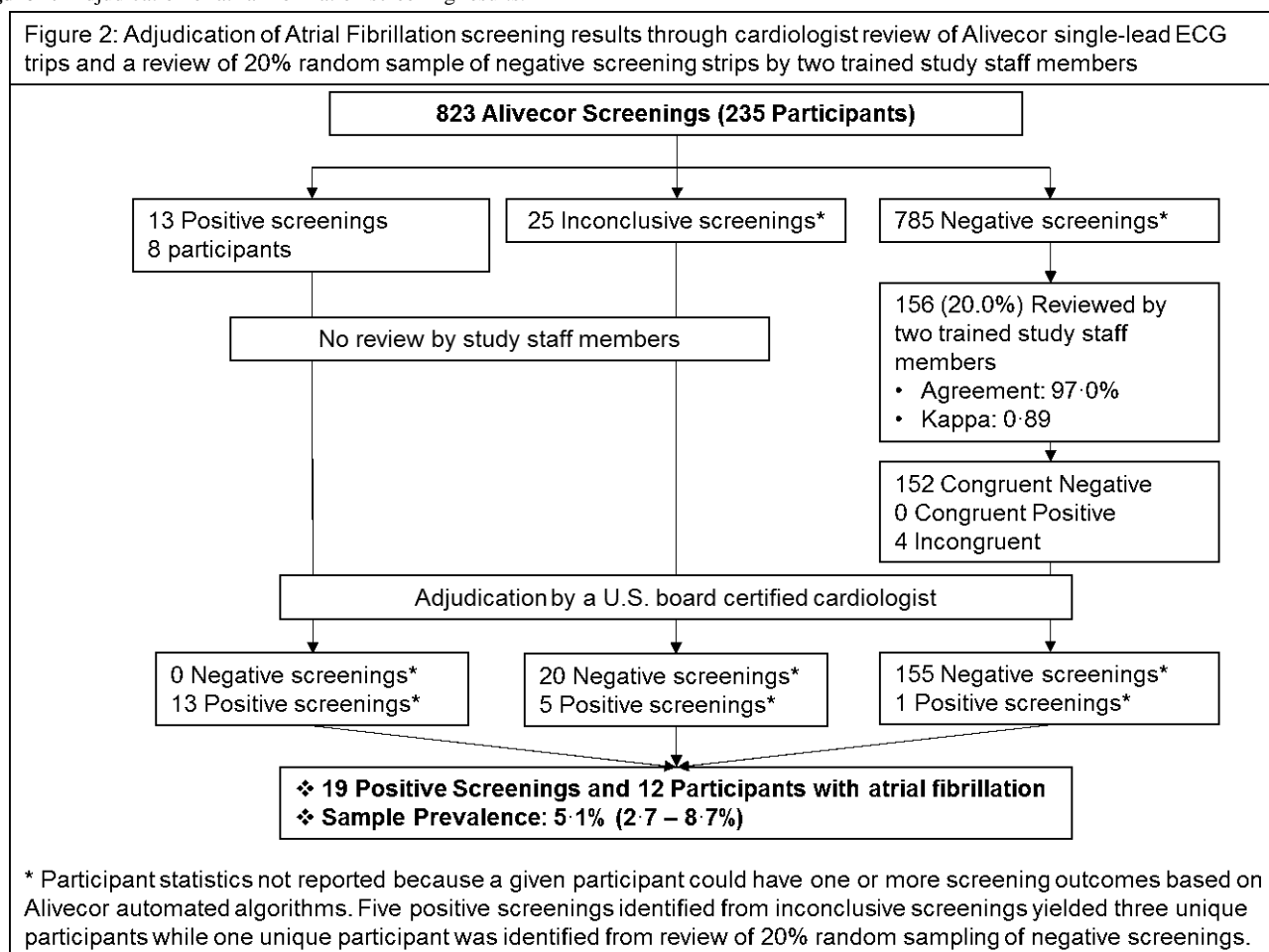
collection of pulse data to develop an automated arrhythmia detection mobile app that can be used in low-resource settings [7,8]. Both Alivecor and pulse data were recorded serially for 2 minutes each on 5 consecutive days over a period of 6 weeks beginning June, 2015. During screening, participants sat cross-legged, resting the smartphone (iPhone 4S) in their lap to stabilize the phone and reduce excess motion that could interfere with the recordings (Figure 1). Additionally, on the day of enrollment, participants responded to a questionnaire that collected information related to their demographic characteristics, lifestyle habits, and past medical history.

The Alivecor device malfunctioned for two weeks, and therefore 120 participants from two villages were not screened for atrial fibrillation using Alivecor and were excluded from this study. Study staff uploaded ECG and pulse check recordings to a secure, Web-accessible Research Electronic Data Capture study database. Because pulse data were collected with the intention of developing an arrhythmia detection app based on the results of this pilot study, our outcome of atrial fibrillation was determined based solely on the ECG results from the FDA-approved Alivecor device. A board-certified cardiologist reviewed all ECG tracings for participants who had abnormal rhythm findings based on the automated Alivecor algorithm (Figure 2). Any participant found to have atrial fibrillation was referred to a study cardiologist located at a regional academic health center. Due to constraints in our available resources, our research staff did not follow up with participants after screening to assess whether any clinical plan was initiated.

A randomly selected 20% subsample of normal ECG tracings were reviewed by two trained study staff members, and discordant readings were adjudicated by the study cardiologist. Thus, a board-certified cardiologist reviewed the ECG tracings of all participants who were determined to have positive screening findings for the presence of atrial fibrillation. The study received institutional review board approval from the University of Massachusetts Medical School and HM Patel Center for Medical Care and Education. Descriptive statistics were utilized to describe the characteristics of study participants. Sociodemographic and comorbid factors were compared across different age groups using Fisher exact tests. Prevalence rates of atrial fibrillation were calculated in a standard manner with accompanying 95% confidence intervals. Given the limited sample size in our pilot investigation and the use of the Alivecor ECG as the source for gold standard measurement, we did not calculate performance measures.

**Figure 1.** Community health worker screening a study participant for atrial fibrillation using a single-lead ECG device.



**Figure 2.** Adjudication of atrial fibrillation screening results.

## Results

Almost two-thirds of study participants were 55 years or older, nearly half were female, one-third did not receive any formal education, and more than one-half were from households earning less than US \$2 per day (Table 1).

Twelve participants screened positive for atrial fibrillation yielding a sample prevalence of 5.1% (95% CI 2.7-8.7) (Figure 2); the characteristics of these individuals are shown in Table 2.

Only one participant had persistent atrial fibrillation throughout all of the screenings; 9 screened positive only once. The cumulative prevalence of atrial fibrillation in this population according to increasing number of screenings is presented in Table 3.

The first screening only identified 7 participants with a positive screen for atrial fibrillation. The remaining 5 participants who screened positive for atrial fibrillation were identified at the fourth screening. A comparison of the 235 participants included in the analyses, with the 120 excluded participants, revealed no meaningful differences between the two groups (see Multimedia Appendix 1 for details).

**Table 1.** Sociodemographic, lifestyle, and health characteristics of 234 participants from rural Gujarat, India, screened for arrhythmias, stratified by age groups.

	N	Age group (%)			P value <sup>a</sup>
		50-55	55-65	65+	
Female <sup>b</sup>	140	71.4	55.0	56.3	.09
<b>Education</b>					.19
None	70	30.7	27.0	35.1	
10 <sup>th</sup> grade or less	129	50.0	59.6	58.4	
More than 10 <sup>th</sup> grade	29	19.4	13.5	6.5	
Works for pay	60	45.8	29.9	9.1	<.001
<b>Daily household income<sup>c</sup></b>					.03
Less than \$1	71	33.9	28.1	31.7	
\$1-\$2	62	14.5	31.5	31.7	
\$2-\$4	53	37.1	20.2	15.2	
More than \$4	44	14.5	20.2	21.5	
Smoking history	37	7.9	14.3	11.3	0.25
Chew tobacco	58	33.3	22.0	21.3	0.03
Hypertension	87	27.0	34.1	48.8	0.02
Diabetes	20	9.5	5.5	11.3	0.37
Hypercholesterolemia	21	4.8	8.8	12.5	0.30

<sup>a</sup>Fisher exact test.

<sup>b</sup>One participant had completed the screening and thus was included in the analyses but did not respond to the questionnaire.

<sup>c</sup>Based on a conservative exchange rate of 1 USD = 60 INR for 2015 calendar year.

**Table 2.** Characteristics of 12 atrial fibrillation positive cases identified by a cardiologist review of single-lead ECG recording.

	Gender	Age	Index positive <sup>a</sup>	# positive <sup>b</sup>	Smoking	Hypertension
1	Female	50-55	3	1/3	No	No
2	Female	55-60	1	1/3	No	No
3	Female	60-65	1	5/5	No	No
4	Female	60-65	1	2/5	No	Yes
5	Female	75-80	1	1/4	No	Yes
6	Female	80-85	1	3/4	No	No
7	Male	50-55	3	1/3	Yes	Yes
8	Male	55-60	1	1/1	Yes	No
9	Male	60-65	1	1/1	No	Yes
10	Male	70-75	4	1/5	Yes	No
11	Male	75-80	3	1/3	No	Yes
12	Male	75-80	4	1/5	No	Yes

<sup>a</sup>Refers to the number of screening when atrial fibrillation was first recognized.

<sup>b</sup>Refers to the total number of positive screenings for a given participant.



**Table 3.** Cumulative prevalence of atrial fibrillation by number of screenings.

Screening number	Cumulative prevalence (95% CI)
1	3.0 (1.2-6.1)
2	3.0 (1.2-6.1)
3	4.3 (2.1-7.7)
4	5.1 (2.7-8.7)
5	5.1 (2.7-8.7)

## Discussion

### Principal Findings

Our study suggests a prevalence of atrial fibrillation in this Indian region (5.1%) that is markedly higher than has been previously reported in India and similar to the prevalence estimates reported in studies of persons from North America and Europe [1,9,10]. This finding is noteworthy and challenges conventional wisdom that individuals of European descent have higher rates of atrial fibrillation than individuals of Asian descent [1].

Current understanding of the global epidemiology of atrial fibrillation is dependent on robust surveillance systems and high quality community-based studies, but there remains a paucity of such investigations outside of North America and Europe, particularly in countries with less developed health systems [10]. A 2012 meta-analysis of community-based screening studies identified only one study from India [10]. That study was conducted in a tribal Himalayan village and found only one case of atrial fibrillation among 984 screened individuals, a prevalence rate of 0.1% [9]. However, 94% of participants in that study were less than 65 years old and thus not representative of the age profile of typical atrial fibrillation patients. A recently published opportunistic screening study of festival attendees reported a slightly higher but still low prevalence of atrial fibrillation (0.5%) among individuals 50 years of age or older [11]. Reasons for the discrepancies between our results and prior studies may include the shortcomings of opportunistic screening efforts involving younger individuals and the use of a single spot-check for atrial fibrillation. Our approach, in contrast to the two prior studies in India, utilized a randomized home-based serial screening of participants aged 50 years and older in order to detect paroxysmal and persistent atrial fibrillation. The higher yield from multiple rhythm checks versus a single check for the detection of paroxysmal atrial fibrillation in the community has been emphasized by other studies [12] and is made evident by our findings. Namely, we observed that out of the 12 participants who screened positive for atrial fibrillation, only one had persistent atrial fibrillation. Moreover, 5 participants who were ultimately found to have paroxysmal atrial fibrillation did not have atrial fibrillation detected during their first screen.

Recently, there has been increased attention in North America and Europe to leverage mobile technology for the screening of

persons with undetected atrial fibrillation [12,13]. The establishment of the National Programme for Prevention and Control of Stroke by the Indian government supports the importance of stroke prevention in India. However, due to the cost of ECG-based screening programs and paucity of trained health professionals in many regions, atrial fibrillation screening has not been possible to date. Our efforts suggest that by engaging community health workers to use novel mobile technologies for arrhythmia monitoring we can screen large numbers of Indians for atrial fibrillation. Our capacity to recruit and serially screen residents of the rural Anand community was strengthened by a long-standing relationship between investigators and community health workers in India.

### Limitations

The findings of our study need to be interpreted with appropriate caution given several concerns and limitations. First, this study is based on a relatively small sample size of 235 participants. Therefore, we have presented information about sample sizes and accompanying 95% confidence intervals to demonstrate the range of possible prevalence estimates consistent with the variability observed in our data. Second, we did not perform a gold standard 12-lead ECG to confirm our positive screening findings. It is important to note, however, that AliveCor devices are FDA-approved and are widely used by cardiologists in diverse clinical settings [14]. Lastly, our cross-sectional study design limits our ability to assess any potential outcomes associated with atrial fibrillation or characterize the clinical presentation of atrial fibrillation in more detail. Therefore, future efforts should explore the feasibility and costs associated with replicating our approach in other environments to define the accuracy of the automated algorithms employed in larger and more diverse cohorts, to create referral mechanisms which can accommodate newly identified patients, to more systematically characterize the clinical presentation of atrial fibrillation (eg, valvular diseases, comorbidities, psychosocial impact), and to demonstrate reduced stroke rates through the primary prevention of stroke in screened populations.

### Conclusions

In conclusion, our study has two important implications: (1) mobile technologies may help overcome resource limitations for screening adults for atrial fibrillation in underserved and low-resource settings and (2) serial screening for atrial fibrillation enhances the ability to identify persons at risk for atrial fibrillation.

## Acknowledgments

We thank Ms Nada Esa, Ms. Dharti Patel, Ms Ami Brahmhatt, and Mr Utsav Patel for supporting data collection efforts. We are grateful to Dr. Jay Himmelstein and the Office of Health Policy and Technology at the University of Massachusetts Medical School for supporting our research.

AS received support from the National Center for Advancing Translational Sciences (TL1-TR001454), and JA received support from the National Institute on Minority Health and Health Disparities (P60-MD006912-05). DDM's time was supported by KL2RR031981, 1R15HL121761-01A1, 1UH2TR000921-02, and 1R01HL126911-01A1 from the National Heart, Lung, and Blood Institute.

The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

## Conflicts of Interest

DDM discloses equity stakes or consulting relationships with Flexcon, Bristol-Myers Squibb, Mobile Sense, ATRIA, and Boston Biomedical Associates. He has also received research funding from Sanofi Aventis, Otsuka Pharmaceuticals, Philips Healthcare, Biotronik, and Pfizer.

## Multimedia Appendix 1

Comparison of sociodemographic, lifestyle, and health characteristics of participants that were excluded versus included in the study.

[[PDF File \(Adobe PDF File\). 14KB - publichealth\\_v2i2e159\\_app1.pdf](#)]

## References

1. Chugh SS, Havmoeller R, Narayanan K, Singh D, Rienstra M, Benjamin EJ, et al. Worldwide epidemiology of atrial fibrillation: a Global Burden of Disease 2010 Study. *Circulation* 2014 Feb 25;129(8):837-847 [FREE Full text] [doi: [10.1161/CIRCULATIONAHA.113.005119](https://doi.org/10.1161/CIRCULATIONAHA.113.005119)] [Medline: [24345399](https://pubmed.ncbi.nlm.nih.gov/24345399/)]
2. Gourie-Devi M. Epidemiology of neurological disorders in India: review of background, prevalence and incidence of epilepsy, stroke, Parkinson's disease and tremors. *Neurol India* 2014;62(6):588-598 [FREE Full text] [doi: [10.4103/0028-3886.149365](https://doi.org/10.4103/0028-3886.149365)] [Medline: [25591669](https://pubmed.ncbi.nlm.nih.gov/25591669/)]
3. GBD 2013 Mortality and Causes of Death Collaborators. Global, regional, and national age-sex specific all-cause and cause-specific mortality for 240 causes of death, 1990-2013: a systematic analysis for the Global Burden of Disease Study 2013. *Lancet* 2015 Jan 10;385(9963):117-171 [FREE Full text] [doi: [10.1016/S0140-6736\(14\)61682-2](https://doi.org/10.1016/S0140-6736(14)61682-2)] [Medline: [25530442](https://pubmed.ncbi.nlm.nih.gov/25530442/)]
4. Deore R, Vora A. Epidemiology and risk factor for atrial fibrillation in India. *J Prev Cardiol* 2014;3(3):507.
5. Binnendijk E, Koren R, Dror DM. Can the rural poor in India afford to treat non-communicable diseases? *Trop Med Int Health* 2012 Nov;17(11):1376-1385 [FREE Full text] [doi: [10.1111/j.1365-3156.2012.03070.x](https://doi.org/10.1111/j.1365-3156.2012.03070.x)] [Medline: [22947207](https://pubmed.ncbi.nlm.nih.gov/22947207/)]
6. Lau JK, Lowres N, Neubeck L, Brieger DB, Sy RW, Galloway CD, et al. iPhone ECG application for community screening to detect silent atrial fibrillation: a novel technology to prevent stroke. *Int J Cardiol* 2013 Apr 30;165(1):193-194. [doi: [10.1016/j.ijcard.2013.01.220](https://doi.org/10.1016/j.ijcard.2013.01.220)] [Medline: [23465249](https://pubmed.ncbi.nlm.nih.gov/23465249/)]
7. McManus D, Chong JW, Soni A, Saczynski JS, Esa N, Napolitano C, et al. PULSE-SMART: pulse-based arrhythmia discrimination using a novel smartphone application. *J Cardiovasc Electrophysiol* 2016 Jan;27(1):51-57. [doi: [10.1111/jce.12842](https://doi.org/10.1111/jce.12842)] [Medline: [26391728](https://pubmed.ncbi.nlm.nih.gov/26391728/)]
8. Soni A, Earon A, Handorf A, Fahey N, Allison J, Chon K. Using mobile technologies to detect arrhythmias in rural India. 2015 Presented at: National Institutes of Health Electr Electron Eng Point Care Technol Healthc; 2015; Bethesda, MD.
9. Kaushal SS, DasGupta DJ, Prashar BS, Bhardwaj AK. Electrocardiographic manifestations of healthy residents of a tribal Himalayan village. *J Assoc Physicians India* 1995 Jan;43(1):15-16. [Medline: [9282630](https://pubmed.ncbi.nlm.nih.gov/9282630/)]
10. Lip GY, Brechin CM, Lane DA. The global burden of atrial fibrillation and stroke: a systematic review of the epidemiology of atrial fibrillation in regions outside North America and Europe. *Chest* 2012 Dec;142(6):1489-1498. [doi: [10.1378/chest.11-2888](https://doi.org/10.1378/chest.11-2888)] [Medline: [22459778](https://pubmed.ncbi.nlm.nih.gov/22459778/)]
11. Saggi D, Sundar G, Nair S, Bhargava V, Lalukota K, Chennapragada S. Prevalence of atrial fibrillation in an urban population in India: the Nagpur pilot study. *Heart Asia* 2016;8:56-59.
12. Svennberg E, Engdahl J, Al-Khalili F, Friberg L, Frykman V, Rosenqvist M. Mass screening for untreated atrial fibrillation: The STROKESTOP study. *Circulation* 2015 Jun 23;131(25):2176-2184 [FREE Full text] [doi: [10.1161/CIRCULATIONAHA.114.014343](https://doi.org/10.1161/CIRCULATIONAHA.114.014343)] [Medline: [25910800](https://pubmed.ncbi.nlm.nih.gov/25910800/)]
13. Lowres N, Neubeck L, Redfern J, Freedman SB. Screening to identify unknown atrial fibrillation: a systematic review. *Thromb Haemost* 2013 Aug;110(2):213-222. [doi: [10.1160/TH13-02-0165](https://doi.org/10.1160/TH13-02-0165)] [Medline: [23595785](https://pubmed.ncbi.nlm.nih.gov/23595785/)]
14. Nguyen HH, Silva JNA. Use of smartphone technology in cardiology. *Trends Cardiovasc Med* 2016;26:376-386.

---

**Abbreviations****ECG:** electrocardiogram**FDA:** US Food and Drug Administration

---

*Edited by G Eysenbach; submitted 19.08.16; peer-reviewed by M Cosentino, Y Wu; comments to author 14.09.16; revised version received 21.09.16; accepted 24.09.16; published 13.10.16.*

*Please cite as:*

Soni A, Earon A, Handorf A, Fahey N, Talati K, Bostrom J, Chon K, Napolitano C, Chin M, Sullivan J, Raithatha S, Goldberg R, Nimbalkar S, Allison J, Thanvi S, McManus D

*High Burden of Unrecognized Atrial Fibrillation in Rural India: An Innovative Community-Based Cross-Sectional Screening Program*  
*JMIR Public Health Surveill* 2016;2(2):e159

URL: <http://publichealth.jmir.org/2016/2/e159/>

doi: [10.2196/publichealth.6517](https://doi.org/10.2196/publichealth.6517)

PMID: [27737818](https://pubmed.ncbi.nlm.nih.gov/27737818/)

©Apurv Soni, Allison Earon, Anna Handorf, Nisha Fahey, Kandarp Talati, John Bostrom, Ki Chon, Craig Napolitano, Michael Chin, John Sullivan, Shyamsundar Raithatha, Robert Goldberg, Somashekhar Nimbalkar, Jeroan Allison, Sunil Thanvi, David McManus. Originally published in *JMIR Public Health and Surveillance* (<http://publichealth.jmir.org>), 13.10.2016. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in *JMIR Public Health and Surveillance*, is properly cited. The complete bibliographic information, a link to the original publication on <http://publichealth.jmir.org>, as well as this copyright and license information must be included.

Original Paper

# Development of an Ecological Momentary Assessment Mobile App for a Low-Literacy, Mexican American Population to Collect Disordered Eating Behaviors

Kay Connelly<sup>1</sup>, PhD; Karen F Stein<sup>2</sup>, PhD, RN; Beenish Chaudry<sup>3</sup>, PhD; Nicole Trabold<sup>4</sup>, PhD, LMSW

<sup>1</sup>School of Informatics and Computing, Indiana University, Bloomington, IN, United States

<sup>2</sup>School of Nursing, University of Rochester, Rochester, NY, United States

<sup>3</sup>Notre Dame, South Bend, IN, United States

<sup>4</sup>Department of Psychiatry, Medical Center, University of Rochester, Rochester, NY, United States

**Corresponding Author:**

Kay Connelly, PhD

School of Informatics and Computing

Indiana University

919 E 10th St

Bloomington, IN, 47408

United States

Phone: 1 812 855 0739

Fax: 1 812 856 5754

Email: [connelly@indiana.edu](mailto:connelly@indiana.edu)

## Abstract

**Background:** Ecological momentary assessment (EMA) is a popular method for understanding population health in which participants report their experiences while in naturally occurring contexts in order to increase the reliability and ecological validity of the collected data (as compared to retrospective recall). EMA studies, however, have relied primarily on text-based questionnaires, effectively eliminating low-literacy populations from the samples.

**Objective:** To provide a case study of design of an EMA mobile app for a low-literacy population. In particular, we present the design process and final design of an EMA mobile app for low literate, Mexican American women to record unhealthy eating and weight control behaviors (UEWCBs).

**Methods:** An iterative, user-centered design process was employed to develop the mobile app. An existing EMA protocol to measure UEWCBs in college-enrolled Mexican American women was used as the starting point for the application. The app utilizes an icon interface, with optional audio prompts, that is culturally sensitive and usable by a low-literacy population. A total of 41 women participated over the course of 4 phases of the design process, which included 2 interview and task-based phases (n=8, n=11), focus groups (n=15), and a 5-day, in situ deployment (n=7).

**Results:** Participants' mental models of UEWCBs differed substantially from prevailing definitions found in the literature, prompting a major reorganization of the app interface. Differences in health literacy and numeracy were better identified with the Newest Vital Sign tool, as compared with the Short Assessment of Health Literacy tool. Participants had difficulty imagining scenarios in the interviews to practice recording a specific UEWCB; instead, usability was best tested in situ. Participants were able to use the EMA mobile app over the course of 5 days to record UEWCBs.

**Conclusions:** Results suggest that the iterative, user-centered design process was essential for designing the app to be made usable by the target population. Simply taking the protocol designed for a higher-literacy population and replacing words with icons and/or audio would have been unsuccessful with this population.

(*JMIR Public Health Surveill* 2016;2(2):e31) doi:[10.2196/publichealth.5511](https://doi.org/10.2196/publichealth.5511)

**KEYWORDS**

feeding and eating disorders; health literacy; socioeconomic status; human-centered computing; user-computer interface; mobile apps

## Introduction

### Ecological Momentary Assessment

Ecological momentary assessment (EMA) is a measurement approach rapidly gaining popularity in public health research because of its utility in capturing complex processes associated with major health problems and their treatment. To date, more than 250 studies have used EMA to investigate critical public health problems ranging from diabetes, hypertension, asthma, tobacco, alcohol and drug use, obesity, and physical inactivity. EMA refers to the collection of individuals' behavioral, emotional, cognitive, and biophysical data as they occur in the natural environment, most frequently using an electronic device. EMA is considered a class of measurement approaches that are unique in that they share a set of three properties: (1) the phenomenon of interest is measured repeatedly over time, (2) measurement occurs within the naturally occurring context, and (3) data are collected as the target event occurs or shortly thereafter [1]. Therefore, EMA eliminates problems associated with retrospective recall of a targeted event (eg, how many cigarettes did you smoke in the last month?) and threats to ecological validity. Rather, EMA provides a detailed picture of the individual's experiences as they naturally occur over time and across situations.

An important advantage of EMA is that both objective and subjective data related to the target event can be measured simultaneously or in close succession. Although some studies collect biomarker data in vivo in conjunction with self-report data, others rely exclusively on self-report data. For example, in several studies of the effects of stress on health outcomes, salivary cortisol was measured several times daily over the course of several days concurrently with self-report measures of affect and stress [2,3]. Other EMA studies focused solely on participant subjective experience with protocols including multiple word-based questionnaires.

Despite the many strengths of EMA, one notable limitation is that data collection often includes complex protocols, and subjective data collection relies on word-based questionnaires that include complex ratings and self-evaluations [4]. As a result, EMA studies often exclude large and particularly vulnerable segments of our population—persons with low health literacy.

Health literacy is defined as “the ability to obtain, process, and understand basic health information and services needed to make appropriate health decisions and follow instructions for treatment.” Health literacy, like general literacy, address three skill sets including the ability to search, comprehend, and use information from continuous prose (brochures, instructions), noncontinuous text in diverse formats (tables, forms), and quantitative tasks. Health literacy is strongly associated with race, level of education, and socioeconomic status (SES) [5,6]. Furthermore, health literacy is affected by cultural and linguistic factors that further complicate comprehension of both printed and verbally presented materials. Estimates suggest approximately 12% of the US population has less than a high school education [7] and more than 36 million adults living in the United States read at or below the third grade level. In terms of health literacy, 49% of persons with less than a high school

diploma demonstrated below basic level healthy literacy [5]. Yet EMA studies rarely report participants' level of education and those studies that do almost universally include samples of highly educated participants. Virtually no studies report literacy or health literacy levels of study participants.

This methodological gap is of great concern for public health researchers as low-literacy populations often experience significant health disparities [8] and attempts to quantify their experiences are limited by the lack of valid and reliable approaches to measurement. This paper presents a case study of how to design an EMA tool for a low-literacy population.

### Contributions

We present a case study of a user-centered, iterative design process with a Western New York rural farming population of Mexican American (MA) women. The goal of the design process was to develop a culturally sensitive, understandable and usable mobile phone app to measure unhealthy eating and weight control behaviors (UEWCBs) in a sample of MA women with low literacy skills. UEWCBs include binge eating, food or calorie restriction, fasting, self-induced vomiting, and use of diet pills, diuretics, and laxatives as a means to lose or control body weight. We focused on UEWCBs in this population for several reasons. First, the prevalence of these behaviors is high in MA women of higher SES, and results of a few studies of MA women living in border states suggest they also are prevalent in low SES MA women [9]. Second, EMA methodology has been used effectively both to identify determinants of UEWCBs and as an approach to treatment in higher SES majority women. Finally, UEWCBs contribute to weight gain and indicators of metabolic syndrome—both of which disproportionately affect MA women. Alcohol and tobacco use are highly correlated with UEWCBs in majority populations [10] and prevalence of use is increasing in populations of MA women [11]. Therefore, items to measure alcohol consumption and tobacco use were included.

The case study started with an EMA protocol designed to measure the behaviors of interest in college-enrolled MA women [12]. Although the parent study did not address diagnosable level eating disorders, definitions of the UEWCBs were based on the *Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition (DSM-IV)* [13]. Problems with a direct adaptation of the EMA and research protocols to a low-literacy population were identified and rectified through 4 phases of the design process meant to (1) identify icons to visually represent the target behaviors that are understandable and culturally relevant to MA women and (2) develop behavior-recording navigation that is intuitive and easy to use.

## Methods

The study protocol was approved by the institutional review board (IRB) of the 2 participating universities. Adopting a user-centered, iterative design process meant that we did not know the exact nature of the phases of the study a priori. The final protocol consisted of 4 user-centered phases (Figure 1), including (1) activities to elicit feedback on icons for the interface, (2) a focus group to clarify the population's

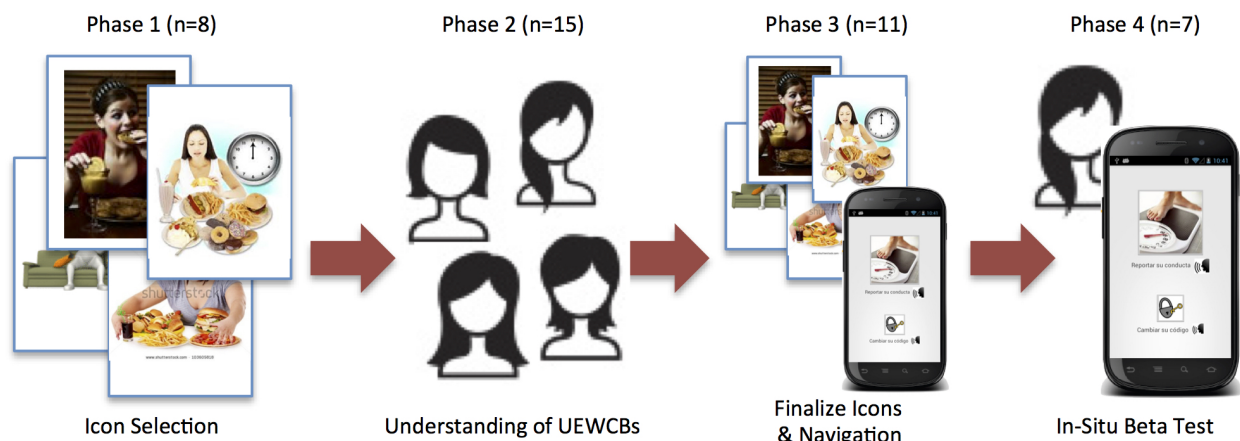


understanding of UEWCBS, (3) activities to finalize icons and test the navigation of the app, and (4) an in situ, 5-day beta test with a fully functional app. All data collection activities were conducted in Spanish. A female member of the community who came to the United States as a migrant farm worker and is widely known by the community as a farm worker rights advocate was employed by our project as a community liaison. In addition, a Spanish-English bilingual woman was employed as data collector.

The community liaison and data collector completed protection of human subjects certification and were trained for data collection by the project manager. Training activities were

completed before each phase and included the subject recruitment protocol, informed consent process, and data collection activities. Although there were slight variations across the phases, generally the community liaison administered the informed consent and collected demographic data. The data collector administered health literacy measures and all study questionnaires. For the beta test, the data collector oriented the participant to the mobile phone app and behavioral recording. The programmer responsible for developing the mobile phone app participated in the first 2 phases to observe participants' mobile phone use competency and ability to navigate early versions of the app.

**Figure 1.** The 4 phases in our final protocol.



## Translations

All study materials, including the informed consents, interview scripts, and measures, were translated into Spanish by a certified translator with fluency in Mexican Spanish. All translations were reviewed by the community liaison and discrepancies or questions were resolved through discussion with the translator. Responses to open-ended questions were translated into English after the sessions by the data collection team.

## Recruitment

The community liaison assumed responsibility for all recruitment activities, including verbal announcements at local churches, day care centers, farm community social activities, and Mexican Consulate on Wheels programs. In addition, the project had IRB approval allowing the community liaison to recruit via home visits. Regardless of the mode of initial contact, the recruitment process included 2 face-to-face visits at the potential participant's home. During the first visit, the community liaison described the project to the potential participant and answered questions. The informed consent was read aloud and the participant was asked to spend a few days talking with her family and thinking over her interest in participating. During the second visit, the participant was asked if she would like to participate and, if she did, an appointment was set for the data collection session.

## Data Collection

Except for the focus groups, all data collection sessions were individual, face-to-face interviews conducted at participants' homes. Both the community liaison and Spanish-speaking data collector conducted all data collection sessions together. Because of issues related to immigration status, names were not recorded on study documents. Informed consent was granted orally with a witness' signature to document agreement. Participants were monetarily compensated for their time, with the amount varying across study phases. Participants were paid in cash (phases 1-3: US \$15, phase 4: US \$55) and a witness' signature documented the payment. Except for one participant who participated in both phase 2 and phase 4, participants were involved in only a single phase of the study.

## Measures

Two measures for health literacy were used to ensure we were reaching our target population. A custom usability questionnaire was used for the final phase to identify potential problems with the app. Psychometrics of the measures was not computed for this study because of the small sample size.

- Short Assessment of Health Literacy–Spanish and English (SAHL-S&E) is an 18-item measure that tests word reading ability and comprehension in the health context. A score of 0-14 was considered inadequate health literacy. SAHL was used for each study. Correlations between SAHL-S and other measures of health literacy (50-item Short Assessment of Health Literacy for Spanish Adults and the

Test of Functional Health Literacy in Adults) and level of education in Spanish-speaking samples were positive and strong. Alpha coefficients for the 18-item English and Spanish versions were .89 and .80, respectively. Results of item response theory indicate that the English and Spanish versions are comparable tests [14].

- Newest Vital Sign (NVS) is a tool to assess health literacy that addresses both reading and numeracy skills [15]. Total scores of 0-1 suggest high likelihood of limited literacy; scores of 2-3 indicate possible limited literacy; scores of 4-6 indicate adequate literacy. Evidence to support construct validity and internal consistency (English:  $\alpha=.76$ ; Spanish:  $\alpha=.69$ ) has been shown [16]. The NVS was used for phases 3 and 4.
- A 43-item usability and context of use questionnaire, modified from [17], was administered at the final interviews for the beta study, which used a working version of the app. Questions focused on factors influencing the recording of behaviors, including overall usability of the app (8) and individual features (icons, 5; reminders, 8; and the camera, 2), learnability and help within the app (5), context of use (11), and how using the app made them feel about their health (4). A total of 31 questions were on a 5-point Likert scale and 12 questions were yes or no answers.

## Analysis

Descriptive statistics were used to analyze data in each phase except phase 2. The sample size for each phase, although appropriate for the specific design activity (eg, Siek et al [18] report a picture card study with  $n=8$ ), is too small to perform inferential analysis beyond averages and counts. In general, if more than one participant had difficulty with an activity, the associated interface component was discussed by the team and improved upon, or further user feedback was elicited. Focus group audiotapes were transcribed and translated. Conventional content analysis was used to analyze the data [19]. Two investigators (KS and NT) participated in the data coding and analysis. Transcripts were read to achieve initial understanding and followed by a line-by-line analysis to derive codes and themes that reflect concepts and thoughts. After all transcripts were coded, the 2 investigators met to discuss any discrepancies in codes, cluster codes into meaningful categories, and discuss potential relationships. During the final phase, means and standard deviations of the usability questionnaire were computed to identify poorly scoring questions, indicating potential problems with the app.

## Phase 1: Icon Selection

A word-based menu-driven interview protocol designed to measure UEWCBS and tobacco and alcohol use in college-enrolled young adult MA women [12] was used as the

starting point for the app development. The first step in the process was to replace word-based representations for the target UEWCBS, tobacco use, and alcohol use behaviors with icons or pictures that were culturally appropriate and easy for our participants to understand. In addition, we sought to add to the protocol measures of the emotional and social context in which the target behaviors occurred. To achieve these aims, a collection of icons and pictures to represent the target behaviors, moods, and contexts were gathered through Web searches completed by the research team. Icons were selected to test a variety of dimensions, such as cartoons versus realistic drawings versus photos, and items shown with and without women. The purpose of phase 1 was to validate the cultural appropriateness and understandability of the selected icons and pictures with the target population.

## Sample

Participants for this phase were 8 women recruited from farming communities in Western New York. All were first generation (eg, born in Mexico) and spoke Spanish as their primary language. The mean age of the sample was 28.3 years (range 21-32 years). Three participants had less than a high school education, 4 completed high school, and 1 completed some college. The mean SAHL score for the sample was 16.1 (range 13-18) with 3 women scoring in the range of low literacy (ie,  $\leq 14$ ).

## Phase 1-Specific Methods

Participants were asked to complete 3 card-sorting tasks to determine meaningfulness, cultural appropriateness, and preferences for the visual images. In task 1, participants were verbally provided definitions of all UEWCBS, tobacco use, and alcohol use behaviors (Table 1, left-hand column), and asked to sort icons into stacks of specific behaviors. Definitions were based on the *DSM-IV* and items from the Eating Disorder Examination [20] but were simplified to improve comprehensibility. Detailed properties of the behaviors were not included in the verbal definition, but plans were for the app to address them with follow-up screens. Once completed, we showed correct icons of each behavior together and asked participants to pick best, worst, and inappropriate icons for each behavior with 2-8 icons of each behavior shown.

In task 2, participants were asked to rate photos and drawings of emotions (happy, sad, angry, frustrated, worried, scared, lonely, and disgusted), locations (car, restaurant, field, store, bed, kitchen, living room, house, and church), activities (resting, working, socializing, exercising, eating, watching television, on the phone, driving, and on computer), and social context (alone, family, friends, coworkers, and crowd) depicted in another set of icons.

**Table 1.** Unhealthy eating and weight control behaviors, alcohol, and tobacco vocabulary and definitions used in the study.

Ten definitions before phase 2: focus groups	Eleven definitions after phase 2: focus groups
Exercise: workout, train, or exercise for the purpose of controlling your weight or shape or to burn calories.	Exercise: physical activities or exercise that you do to lose weight or avoid gaining weight.
Restricting: purposefully limiting the amount of food or calories that you ate or dieting to control your shape or weight. This includes the use of lower calorie meal replacement products such as Slim-Fast.	Changing what you eat or how you eat (diet): changing what or how much you ate to lose or control your weight or deal with your worries about being or getting fat. This would include cutting down how much you ate, cutting out certain foods, skipping meals, and not eating anything for a period of time.
Fasting: taking in no food for a period of several hours to control your shape and weight.	
Binge eating: you've eaten so much food in a short period of time that you would be embarrassed if others saw you, and during these times when you ate this way you feel you couldn't stop eating or control what or how much you were eating.	Eat and eat: overeating, eating a large quantity of food that makes you feel guilty or embarrassed, accompanied by feelings of being out of control.
Diet pills and appetite suppressants: use of pills or other substances for the purposes of burning calories, altering your appetite or metabolism for the purpose of controlling your shape and weight.	Pills: pills that you take because of your worries about your weight and shape. Pills that cause you to get rid of food, reduce your appetite, lose weight, or avoid gaining weight.
Laxatives: taking pills or other types of substances to make your bowels move, to get rid of food or calories to control your shape or weight. Examples: pills, liquids, suppositories, enemas. Common brands may include but are not limited to Ex-Lax, Docusate Sodium, Metamucil, Milk of Magnesia, Citrucel, Fleet.	Powders: powders used to get rid of food, reduce your appetite, or cause you to lose weight or avoid gaining weight.
Diuretics: water pills or other substances for the purpose of reducing the amount of water in your body to control your weight or shape. This includes herbal pills and teas taken to reduce water weight.	Tea: tea that you drink to get rid of food, reduce your appetite, or cause you to lose weight.
	Drinks: drinks that are used to help you lose weight, sometimes they are taken instead of eating a meal; other drinks may be used to reduce your appetite or help you get rid of food.
	Drops: liquid drops that are put into food or water to help you get rid of food, reduce your appetite, or cause you to lose weight.
Vomiting: throw up to control your shape or weight or to get rid of food.	Vomiting: making yourself throw up to get rid of food or avoid gaining weight.
Smoking: use of cigarettes.	Smoking: use of cigarettes or tobacco products.
Alcohol: use of any alcoholic beverages.	Alcohol: use of any alcoholic beverages.

Task 3 involved two behaviors of interest that require the recording of time (binge eating and exercise). We showed participants two, 6-icon sets depicting various times. Participants were asked to select an icon from each set to report 50 minutes of exercise.

After the 3 card-based tasks were completed, participants were asked about the desirability of audio help within the app.

## Results

Performance on task 1 and translated audiotapes of questions during the task revealed that participants (1) did not understand the definitions of the UEWCBS and (2) often did not understand the link between the definitions provided and the icons. In particular, all of the participants had difficulty distinguishing between fasting and restricting images, with the audiotapes revealing they considered fasting to be a (extreme) form of restricting. Likewise, 7 of the participants had difficulty distinguishing images of laxatives, diuretics, and appetite suppressants, with the audiotapes revealing that most simply thought of products that would help them lose weight. Furthermore, images were not always understood as

representations of UEWCBS. For example, icons to represent diet pills were understood as a general picture of pills. Similarly, icons to represent self-induced vomiting to control weight were interpreted as "being sick." Finally, the concept of binge eating was unfamiliar and no suitable translation was identified.

In task 2, participants accurately identified photos and drawings of all 8 emotions and preferred photos of humans depicting the emotions compared with line drawings. Participants had considerable difficulty correctly identifying pictures related to location, activity, and social context.

In task 3, five participants correctly responded to the time task using the clock image and 7 reported that the clock image was easier to read.

Finally, all 8 participants indicated they would like the ability to play the text shown on the pictures aloud. However, 4 of 8 indicated they would not want audio to be played when other people were around.

## Discussion

There were two primary findings that led us to change our research protocol. The first was related to our study methodology: the SAHL measure of health literacy was not adequate to capture the range of cognitive functions required by the study tasks. An adequate literacy score using SAHL simply means a participant can read health vocabulary out loud and recognize word meaning; however, SAHL does not test more complex comprehension and application of concepts. Five of our 8 participants had adequate health literacy according to their SAHL scores, yet all of our participants struggled with understanding UEWCB definitions and 3 had difficulty interpreting time values. An investigation of the literature led us to add NVS to our future protocols, as NVS can be used to measure more complex manipulation of information. Our hope was that NVS would be a more complete measure of health literacy skills required to use a mobile app.

The other important finding indicated that making the EMA app appropriate for a low-literacy and low-acculturated population was not as easy as translating the existing protocol from text to icons, because the language and mental models of the targeted behaviors differed between our population and a higher-literacy MA population. Indeed, behaviors identified and defined in the original protocol suitable for high-literacy MA women were unfamiliar to our target population. Furthermore, the study showed that icons and pictures related to social context were unclear and confusing. These results made us go back to our target group to determine how they think about weight control behaviors so we could accurately capture that in an EMA app.

## Phase 2: Understanding UEWCBs

To gain an understanding of the conceptualization and language of UEWCBs of our target population, 2 focus groups were added to the study protocol. The goals of these focus groups were to (1) describe the weight and shape concerns of Mexican women living in Western rural New York, (2) describe behaviors employed to lose or control weight, and (3) identify the terminology used for these behaviors.

Purposeful recruitment occurred through bilingual community leaders who are well known in the community for being advocates for the local Mexican population (eg, project community liaison and a nutritionist who works at a local health center that services the farm worker community). The community leaders identified potential participants based on eligibility criteria and their knowledge of community members. Potential participants were provided with information flyers, and within 24-48 hours, one of the bilingual study team members contacted them to answer any question and determine interest. Interested women were informed of the date, time, and location of one of the focus groups.

## Sample

Eligibility criteria included MA women between the ages of 18 and 45 years. Most participants immigrated or migrated to the United States, primarily spoke Spanish, and worked in agriculture directly on farms or factories related to the

agriculture industry. All participants lived in rural communities located in Western and Central New York. Because assessing participant literacy levels would have necessitated individual administration of a written questionnaire that would have significantly increased the duration of the meeting time and burden of participation, we did not collect specific demographic information or conduct any literacy or cognitive tests. Each focus group was composed of 7-8 women.

## Methods

The project community liaison and a bilingual nutritionist who also lives and works in the participants' community facilitated the focus groups. The principal investigator (KS) and project coordinator (NT) participated as observers. The focus groups were held in community locations and free childcare was offered. The informed consent process was completed in the group setting with verbal consent because of immigration issues. A semistructured interview format was followed for the group meetings with probes to encourage detailed responses. Both groups were conducted in Spanish and lasted approximately 90 minutes. Both were audiotaped.

## Results

The detailed findings of the focus groups are reported elsewhere (manuscript under review). Here, we summarize the findings most relevant to the design of the app. We found that protocol labels for behaviors were not meaningful to the target population. First, participants recognized behaviors involving a product in terms of product form rather than product outcome, for example, a "diet pill" as opposed to a "laxative." Second, participants tended to describe the actual eating and weight control strategy rather than label it, for example, "eat and eat" versus "binge eat" and "eat smaller portions" versus "restrict." The language of behaviors that emerged was as follows: change what you eat, eat and eat, powders, drops, teas, shakes, pills, vomit, and exercise. The right-hand side of [Table 1](#) provides the UEWCB labels and definitions used for the remaining phases included in this project.

## Discussion

In order to make the EMA app usable by participants, the app must employ their language and concepts for behaviors. However, researchers need to be able to label each participant-recorded behavior with the original protocol terminology. For example, it is not enough for researchers to know that women take pills to lose or maintain their weight. Researchers are interested in the types of pills they take (eg, is it a diuretic, appetite suppressant, or laxative?). This has implications for public health interventions. Thus, the app has to record enough information about the product being consumed (or behavior being engaged in) to map it to the original terminology ([Table 1](#), left column). This has implications for the app design, requiring users to record additional information after a high-level behavior is selected.

With this new conceptualization and vocabulary identified, we added another picture card portion to the originally planned navigation phase. This picture card portion was used to identify appropriate icons for the redefined and newly identified behaviors.



### ***Phase 3: Finalize Icons and Navigation***

A new collection of images depicting UEWCBS and context was gathered by our research team and tested in phase 3. Participants also were shown an early version of the app on a mobile phone and asked to navigate the app to record specific UEWCBS. This early version used a linear navigation and the Android hardware Home button, as has been shown to be usable by other low-literacy populations [17].

#### **Sample**

Participants were 11 women of Mexican origin (all first generation and Spanish speaking). The mean age was 29.1 years (range 20-36 years). The average years of education was 8.5 with 3 of the 11 women completing high school and 3 completing grammar school only. The mean SAHL score for the sample was 14.2 (range 8-17) with 4 women scoring in the range of low literacy on this measure. The NVS measure of health literacy was added to our protocol for this phase. The NVS scores ranged from 0 to 3 with a mean score of 1, suggesting that 7 participants had high likelihood of low literacy and 4 participants had possible low literacy.

#### **Methods**

Part 1 of this phase was composed of tasks 1 and 2 described under phase 1. Revised definitions of UEWCBS based on focus group results and corresponding photos and drawings gathered from the Web were used as stimuli for task 1. Given the difficulty participants had understanding the images for context, new photos and drawing were tested. In part 2 of this phase, participants were asked to record behaviors on a preliminary version of the app installed on the mobile phone. Brief scenarios of UEWCBS were read aloud, and participants were asked to record the behavior described on the mobile phone app.

#### **Results**

Results for the picture card portion of this third phase were much improved. Conceptions of UEWCBS and related photo images were understood with 7 of the concepts being unanimous and 5 of the concepts having a single woman disagree that the final selected photo was a good representation of the concept. However, meaning of photos and images to denote context continued to be problematic for the majority of participants. Furthermore, although participants could correctly identify all of the icons depicting emotions, in both this phase and the first phase, participants only ever indicated that they personally felt positive emotions.

Results from the navigation portion of our third phase indicate that the task of recording behaviors on the app was difficult for our population. We determined this was at least in part because of the way we structured the tasks. Initially, the scenarios were written in the first person asking the participant to imagine that they just engaged in a target UEWCBS. Participants had difficulty with the “imagining task” and could not transfer the example to the app. In some cases, participants would argue that they would never perform the behavior we just had them imagine (eg, “I would never vomit on purpose!”), whereas in other cases participants had difficulty shifting between internal and external states over time (eg, reporting on an emotion they

felt immediately before a behavior, then subsequently reporting their physical location at the time of the behavior).

#### **Discussion**

Although participants endorsed experiencing negative emotions in relation to their weight, shape, and eating behaviors in the focus groups, participants in phases 1 and 3 only indicated having positive emotions. It is unclear why this discrepancy exists, but is likely a result of the methodology. Phases 1 and 3 were conducted with a single participant and 2 researchers. The formal nature of the interviews and the perceived power imbalance between researchers and participant could have played a role. Similarly, participants could have wanted researchers to perceive them positively or simply could have been much more reserved in that setting. With the focus groups, participants interacted with each other much more than with the research team. It was a more relaxed atmosphere, with a lot of laughing and storytelling. Further research is needed to understand the exact ways in which the two methods affect how participants describe UEWCBS and its context, but other researchers should be aware of this difference when studying similar sensitive health behaviors.

On the basis of results of the navigation portion of this phase, a decision was made by the research team to narrow the scope of the app to focus only on recording UEWCBS, tobacco use, and alcohol use behaviors. We determined that recognition and recording of mood and context just before the behavior was complex and would render the app too demanding to be reliably used. We also decided that answering the questions around the navigability of the app would have to be done in situ, when participants were actually engaging in the behaviors of interest.

### ***Phase 4: In Situ Beta Test***

We conducted a beta test with the target population to assess the usability of the app. We were most interested in determining if recording their UEWCBS with the mobile app was intuitive, if the reminders were appropriate, and if the participants were able to use the app throughout their normal day.

#### **Sample**

Seven MA women were recruited. Five participants were born in Mexico but had been in the United States for an average of 14.6 years. The remaining 2 were born and raised in the United States. The average age was 28.8 years (SD 5.72 years). Four participants were employed at the time of the study; 1 packaged vegetables, 1 cultivated crops, 1 assisted in a day care center, and 1 worked as a certified nurse assistant. All participants except one had 2-3 children in the age range of 1 to 14 years. Four spoke English but the main language of communication with friends and family was Spanish. Two also spoke another Mexican dialect—Mestizo.

All Mexican-born participants had completed schooling in Mexico—2 had completed high school, the remaining 3 had only attended until middle school. All participants owned a mobile phone and all had used some mobile apps and mobile Internet. The most popular apps were “YouTube” and “Facebook.” One participant had used health and wellness mobile apps such as Weight Watchers and Healthy Women. On



the basis of the NVS, 5 of our 7 participants had low literacy skills.

## Methods

Participants received training to perform basic phone operations such as turning it on and off, use the back key, and so on. They were then trained to use the mobile EMA app. More specifically, participants learned to enter various behaviors, listen to instructions, change their passwords, make audio recordings, and take photographs. A picture-based manual was used to reinforce the topics covered during training. On average, a training session lasted for 1.5 hours.

Participants were then given the phone (and the training manual) to record their behaviors. Since the beta study aimed at testing the usability of the app, we deployed it for only 5 days. To encourage women to use the study phone, all phones were network activated and participants could use the phone and texting capabilities for personal use.

Participants were instructed to record behaviors each time they occurred. In addition, they also had to respond to 3 app-generated signals each day. At the completion of the study, we met with participants to collect the phones and administer questionnaires on usability and usage of the app. Participants were paid US \$3 per signal response (3 signals/day × 5 days with maximum total US \$45). Signal response did not mean a participant had to record a behavior—they could indicate that no behavior had occurred. In addition, they received US \$10 for completing the post-study usability questionnaire and a

semistructured interview to elicit their feedback on the app (total compensation up to US \$55).

## Results

The EMA data from the beta study indicate that participants were able to use the app to record their UEWCBS. The 7 participants recorded a total of 135 behaviors. All behaviors were recorded by participants from 2 to 27 times. The most complex behavior to record was binge eating, which was selected a total of 16 times. Of these, 11 of the entries had complete follow-up questions (ie, duration of behavior, feelings of embarrassment, loss of control, and context of behavior). Of the 5 instances where the follow-up questions were not completed, participants recorded other behaviors during those sessions, indicating that participants likely selected binge eating accidentally, then returned to the main menu to select the correct behavior.

The usability questionnaire indicates that participants found the app easy to use, with none of the Likert items averaging below a 3 (on a 5-point scale). Table 2 lists the mean scores for the 7 items with a standard deviation more than 1, indicating there was disagreement among participants. Upon closer examination, the first 4 items had all been ranked neutral or above, except for a single participant. Transcriptions of the sessions with this participant indicated she had significant difficulty with the app. The last 3 items had more significant differences, with 2-3 participants rating the question negatively. This indicates that 3 participants selected the wrong icons when navigating the app, 2 did not want others to know they were using the app, and 3 did not often take pictures of products (eg, laxatives).

**Table 2.** Usability items with SD>1 for beta study.

Likert usability questions with SD > 1	Average	SD
31. When I began using the cell phone, I had no idea what I was doing. <sup>a</sup>	4.00	1.15
7. Even after I used the application for several days, I needed a lot of help to use it. <sup>a</sup>	3.86	1.07
19. The phone application gave too many reminders. <sup>a</sup>	3.86	1.07
27. I remembered to take the phone with me wherever I went.	3.86	1.07
10. I picked the wrong picture many times. <sup>a</sup>	3.14	1.07
28. I didn't want people to know I was using the application. <sup>a</sup>	3.14	1.21
16. I often took pictures of products.	3.00	1.67

<sup>a</sup> Denoted reverse coding. Items have been reverse coded so a high score reflects high disagreement.

The yes or no questions in the usability questionnaire were all unanimously positive, except for 3, all relating to context of use. Five participants indicated they did not use the app in front of friends, 6 did not use it in front of strangers, and 4 did not use it at work.

Finally, transcripts of the exit interview indicate that participants were generally unable to register their behaviors right at the time they took place, electing to record in response to reminders or at fixed times during the day. They found the reminders useful and appreciated the 30-minute response window. Some participants also requested being able to select diet throughout the day, and not just at the end of day reminder.

## Discussion

On the basis of participant feedback, we modified the final app in two primary ways. First, we allowed participants to select time periods for reminders during the initial application setup with researchers. This would reduce the possibility that the reminders would occur during inconvenient times, such as when they were working. Second, the app initially asked about dieting behaviors only at the end of the day reminder, because prior research had simply examined if a person had dieted for a particular day or not [12]. However, participants wanted to record every time they consciously performed a dieting behavior (eg, skipping lunch was a separate action from reducing a portion during dinner), and not just wait until the end of the

day. Thus, the final app design includes dieting behaviors in the main menu so women can record them at any time.

### Final Ecological Momentary Assessment App

The EMA mobile app was developed for Android mobile phones. Figures 2-5 show the interface. The landing page (Figure 2, part a) has an option to start a recording or to change their password to enter the app. On the recording page (Figure 2, parts b and c), the user can select from 11 options, which include the UEWCBs of interest, as well as alcohol and tobacco use, as described in the right-hand column of Table 1. Once a recording is complete, the user is returned to the landing page and an

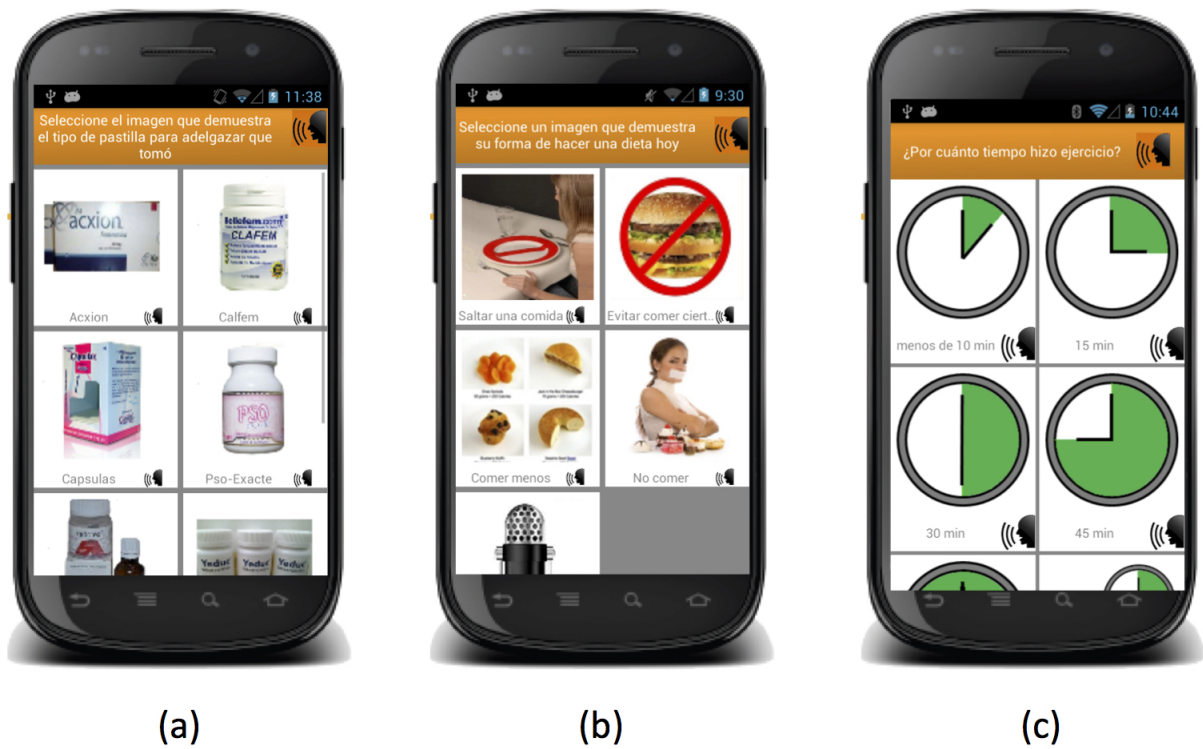
acknowledgment of the recording is briefly shown on the screen (Figure 2, part d).

Of all of the UEWCBs, only vomiting does not have any follow-up screens. The other behaviors prompt the user to select the specific product they consumed (eg, Figure 3, part a), record the kind of dieting behavior they engaged in (Figure 3, part b), the amount of time they engaged in physical activity (Figure 3, part c), the number of cigarettes or type and number of alcoholic drinks they consumed (Figure 4, parts a to c), or answer questions to determine if they engaged in an eating episode that met the criteria for a binge episode (Figure 5). The interface was primarily made up of large pictures with short Spanish labels. If a participant was unable to read the text, every label could be played out loud by pressing the speaker icon next to it.

**Figure 2.** Screenshots of the interface: part (a) is the landing page and parts (b and c) are a selection of 11 behaviors to record (user must scroll to see items at the bottom). From top to bottom, left to right: changing what you eat or how you eat, exercise, eat and eat, pills, powders, teas, drinks, drops, vomiting, smoking to lose weight, and drinking. Part (d) shows acknowledgment that recording was successful.



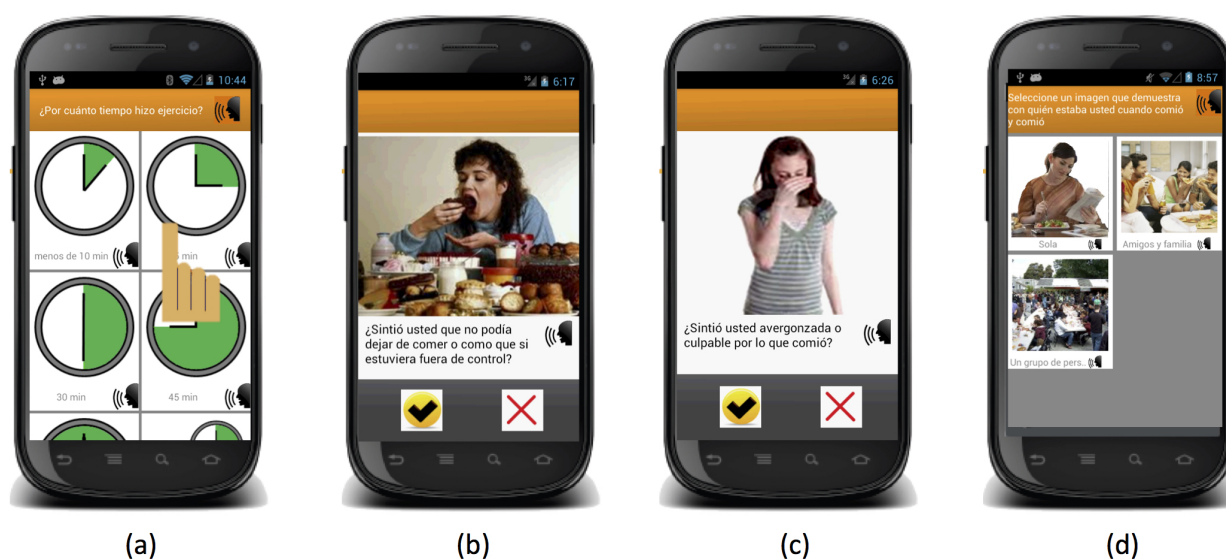
**Figure 3.** Screenshots of the interface: part (a) is a selection of pill products, (b) different types of dieting behaviors, and (c) time selection when recording exercise.



**Figure 4.** Screenshots of the interface: (a) recording the number of whole and partial cigarettes smoked; (b) different types of alcoholic drinks—mixed, wine, beer, and; (c) number of drinks when wine is being recorded.



**Figure 5.** Screens to determine a binge eating episode, including (a) time interval in which eating occurred, (b) feeling a loss of control, (c) feeling embarrassed, and (d) who the person was with at the time of the episode.



## Discussion

Table 3 summarizes the major findings and their implications throughout all 4 phases. Here, we discuss the most significant findings that either changed the design of the EMA app or changed the way in which we performed the research with this population.

The iterative design process identified two significant findings that influenced the design of the app. First, low-literacy and low-aculturated MA women think differently about UEWCBS than researchers in three ways: (1) they used descriptions of the behavior instead of more formal terminology (eg, “eat and eat” instead of “binge eating”), (2) they did not differentiate between nuances of the same behavior (eg, restricting calories and fasting), and (3) they focus on form factor of a diet product instead of the underlying mechanism (eg, a pill, powder, tea, drink, or drops instead of appetite suppressant, laxative, or diuretic). The interface necessarily had to reflect the way the population thought about their behaviors in order to allow them to accurately record them. This required the team to make a mapping from the behavior or product selected by the participant (eg, a specific pill) to the UEWCBS of interest (eg, a laxative).

Second, the original protocol collected each behavior’s emotional, social, and locational context. However, from the initial user studies, we learned that women were lost when they

could not find an icon matching their mental models or experiences of behaviors. We therefore simplified the protocol to collect type and description of behaviors only. Emotional and social contexts were required only for “eat and eat” (which corresponds to the formal definition of binge eating).

The design process also prompted us to refine our study protocols in two significant ways. First, we determined that a standard measure of health literacy (SAHL), which measures health vocabulary, was insufficient to differentiate between participants’ literacy and numeracy skills. The NVS measures the ability to translate health information to practice.

Second, performing in-lab user studies in which participants are asked to “imagine” a specific scenario before using the app was too confusing for our target population. Instead, studying the navigation of the app had to occur in situ.

Without an iterative (and flexible) user design process, the EMA app would likely have been too complex for our target population to use. The next step in our research is to use the EMA app with a much larger sample ( $n=60$ ) of our target population for 3 weeks each, in order to document the prevalence of UEWCBS in this population. This will enable public health researchers to determine the types of interventions that are needed by low-literacy and low-aculturated MA women.



**Table 3.** Summary of major findings and their implications for each phase during the iterative, user-centered design process.

Phase	Finding	Implication/action item
<b>Phase 1: picture card study with 3 tasks</b>	Participants did not understand UEWCB <sup>a</sup> definitions used by researchers.	Could not simply translate existing EMA <sup>b</sup> protocol from text to pictures. Added focus groups to elicit their understanding of and language for UEWCBs.
	Participants accurately identified emotion icons but did not endorse negative emotions.	
	Participants had difficulty with location, activity, and social context icons.	Found more icons to test.
	Most participants could use the clock image for reporting length of time.	Include clock as part of EMA interface for recording binge eating and exercise.
	All participants wanted the ability to play audio to read text out loud.	Include option to play all text in audio.
	SAHL <sup>c</sup> was not adequate for identifying cognitive abilities that affect participant ability to use a mobile app.	Added NVS <sup>d</sup> to our protocol.
<b>Phase 2: two focus groups</b>	New language for some UEWCBs.	Used terminology familiar to participants in EMA app and training materials.
	Focus on form rather than function of UEWCB products.	Matched interface to mental model of participants. Had to expand application to gather enough information to map product form to function.
<b>Phase 3: picture card and navigation study</b>	Icons for new UEWCB terminology agreed upon.	Use appropriate icons in EMA app.
	Contextual icons still problematic.	Remove contextual questions <i>except</i> for binge eating.
	“Imagining tasks” were problematic.	Fully evaluate usability in situ.
	Participants only endorsed positive emotions, in contrast to focus groups.	Keep in mind that participants may be reluctant to record certain emotions.
<b>Phase 4: in situ beta study</b>	Participants could and did use the app to record a variety of UEWCBs in a variety of contexts.	
	Participants appreciated the reminders to record.	Retain reminders.
	Participants did not want to be disturbed during work.	Include customizable time periods for 3 reminders.
	Participants wanted to record every time they engaged in a dieting behavior, not just once a day.	Including dieting and restricting behaviors in main behavior page, not just end of day reminders.

<sup>a</sup> UEWCB: unhealthy eating and weight control behavior.

<sup>b</sup> EMA: ecological momentary assessment.

<sup>c</sup> SAHL: Short Assessment of Health Literacy.

<sup>d</sup> NVS: Newest Vital Sign.

## Limitations

Our design process was exclusively with MA women. Their UEWCB concerns and mental models are likely very different from other low-SES women, making many of the results very specific to this population. Because of privacy concerns and participant burden, demographic information was not collected for the focus groups in phase 2, although we are reasonably confident that the women do meet our inclusion criteria given the recruitment methods. Although the sample size was adequate for making design decisions, it was too small to perform statistical analysis for reporting, preventing us from correlating performance with literacy levels. Our subsequent field deployment will enable such analysis.

## Comparison With Prior Work

Most ESM studies assume adequate literacy and present the ESM questions in text form (eg, [2-4]). Ours is the first visual

ESM implementation designed specifically for low literacy. Freedman and colleagues [21] performed an ESM study that was sensitive to the issue of low literacy, implementing their protocol through a phone-based menu system. They found that participants became frustrated with having to listen to the phone menu every time they called in and attempted to memorize the numbers associated with specific choices in an effort to skip the prompts. Because our system was visual, our participants could efficiently select the behavior they needed to record and did not experience the frustration of having to listen to a series of choices before finding the relevant one for a given recording session.

## Conclusions

There is a striking absence in the literature of using ESM apps to understand health behaviors of low-literacy populations. This paper provides evidence that an ESM app can be developed for those with low literacy. Emphasis must be placed on the



interface mimicking how the population thinks about the behaviors, so they can easily record the behaviors of interest.

Although we were successful in designing an app for women to record a variety of UEWCBS, our attempts to record contextual information fell short. In particular, our participants could record the length of time for a particular behavior but had difficulty recording activity, location, and social context. Further

research is required to determine the best way to obtain this type of context information from this population. A potential solution may be to automatically collect probable context with technology using Global Positioning System and state-of-the-art activity recognition algorithms. This would lessen the burden on participants and provide researchers much needed contextual information.

---

## Acknowledgments

This work was funded by a grant from the National Institute of Nursing Research: 415831-G (Principal Investigator KS).

---

## Authors' Contributions

All authors were involved in conducting the research. KC and KS collaboratively wrote the manuscript. NT and BC participated in data collection and analysis, and reviewed and provided comments on the manuscript. In addition, BC implemented the mobile app and provided technical support during the beta study.

---

## Conflicts of Interest

None declared.

---

## References

1. Shiffman S, Stone AA, Hufford MR. Ecological momentary assessment. *Annu Rev Clin Psychol* 2008;4:1-32. [Medline: [18509902](#)]
2. Entringer S, Buss C, Andersen J, Chicz-DeMet A, Wadhwa PD. Ecological momentary assessment of maternal cortisol profiles over a multiple-day period predicts the length of human gestation. *Psychosom Med* 2011;73(6):469-474 [FREE Full text] [doi: [10.1097/PSY.0b013e31821fbf9a](#)] [Medline: [21700714](#)]
3. Steptoe A, Gibson EL, Hamer M, Wardle J. Neuroendocrine and cardiovascular correlates of positive affect measured by ecological momentary assessment and by questionnaire. *Psychoneuroendocrinology* 2007 Jan;32(1):56-64. [doi: [10.1016/j.psyneuen.2006.10.001](#)] [Medline: [17157442](#)]
4. Wilson SJ, Smyth JM, MacLean RR. Integrating ecological momentary assessment and functional brain imaging methods: new avenues for studying and treating tobacco dependence. *Nicotine Tob Res* 2014 May;16 Suppl 2:S102-S110 [FREE Full text] [doi: [10.1093/ntr/ntt129](#)] [Medline: [24132411](#)]
5. Kutner M, Greenberg E, Jin Y, Paulsen C. ERIC. Washington, DC: National Center for Education Statistics; 2006. The Health Literacy of America's Adults: Results From the 2003 National Assessment of Adult Literacy (NCES 2006-483) URL: <http://files.eric.ed.gov/fulltext/ED493284.pdf> [accessed 2016-06-15] [WebCite Cache ID 6iHeCIQaA]
6. American Speech-Language-Hearing Association. 2016. URL: <http://www.asha.org/slp/healthliteracy/> [accessed 2016-01-08] [WebCite Cache ID 6eNygVggb]
7. Bureau USC. 2016. URL: [https://www.census.gov/newsroom/cspan/educ/educ\\_attain\\_slides.pdf](https://www.census.gov/newsroom/cspan/educ/educ_attain_slides.pdf) [accessed 2016-01-08] [WebCite Cache ID 6eNz3LcgI]
8. Schalick LM, Hadden WC, Pamuk E, Navarro V, Pappas G. The widening gap in death rates among income groups in the United States from 1967 to 1986. *Int J Health Serv* 2000;30(1):13-26. [Medline: [10707297](#)]
9. Suarez L, Felkner M, Brender JD, Canfield MA. Dieting to lose weight and occurrence of neural tube defects in offspring of Mexican-American women. *Matern Child Health J* 2012 May;16(4):844-849. [doi: [10.1007/s10995-011-0806-9](#)] [Medline: [21512779](#)]
10. Saules KK, Pomerleau CS, Snedecor SM, Mehringer AM, Shadle MB, Kurth C, et al. Relationship of onset of cigarette smoking during college to alcohol use, dieting concerns, and depressed mood: results from the Young Women's Health Survey. *Addict Behav* 2004 Jul;29(5):893-899. [doi: [10.1016/j.addbeh.2004.02.015](#)] [Medline: [15219333](#)]
11. Hardie TL, Polek C, Garcia V, González L, Welsh M. Drinks per day in women of Mexican origin: does birth place matter? *J Addict Nurs* 2013;24(3):173-179. [doi: [10.1097/JAN.0b013e3182a4cc0a](#)] [Medline: [24621547](#)]
12. Stein KF, Chen DD, Corte C, Keller C, Trabold N. Disordered eating behaviors in young adult Mexican American women: prevalence and associations with health risks. *Eat Behav* 2013 Dec;14(4):476-483. [doi: [10.1016/j.eatbeh.2013.08.001](#)] [Medline: [24183140](#)]
13. American Psychiatric Association. Diagnostic and statistical manual of mental disorders: DSM-IV. Washington, DC: American Psychiatric Association; 2000.
14. Lee SY, Stucky BD, Lee JY, Rozier RG, Bender DE. Short Assessment of Health Literacy-Spanish and English: a comparable test of health literacy for Spanish and English speakers. *Health Serv Res* 2010 Aug;45(4):1105-1120 [FREE Full text] [doi: [10.1111/j.1475-6773.2010.01119.x](#)] [Medline: [20500222](#)]

15. Weiss BD, Mays MZ, Martz W, Castro KM, DeWalt DA, Pignone MP, et al. Quick assessment of literacy in primary care: the newest vital sign. *Ann Fam Med* 2005;3(6):514-522 [FREE Full text] [doi: [10.1370/afm.405](https://doi.org/10.1370/afm.405)] [Medline: [16338915](https://pubmed.ncbi.nlm.nih.gov/16338915/)]
16. Weiss BD, Mays MZ, Martz W, Castro KM, DeWalt DA, Pignone MP, et al. Quick assessment of literacy in primary care: the newest vital sign. *Ann Fam Med* 2005;3(6):514-522 [FREE Full text] [doi: [10.1370/afm.405](https://doi.org/10.1370/afm.405)] [Medline: [16338915](https://pubmed.ncbi.nlm.nih.gov/16338915/)]
17. Connelly K, Siek KA, Chaudry B, Jones J, Astroth K, Welch JL. An offline mobile nutrition monitoring intervention for varying-literacy patients receiving hemodialysis: a pilot study examining usage and usability. *J Am Med Inform Assoc* 2012;19(5):705-712 [FREE Full text] [doi: [10.1136/amiainl-2011-000732](https://doi.org/10.1136/amiainl-2011-000732)] [Medline: [22582206](https://pubmed.ncbi.nlm.nih.gov/22582206/)]
18. Siek K, Connelly K, Rogers Y. Pride and prejudice: learning how chronically ill people think about food. In: Proceedings of the SIGCHI conference on human factors in computing systems.: ACM; 2006 Presented at: SIGCHI conference on human factors in computing systems; April 24-27, 2006; Quebec, Canada p. 947-950.
19. Hsieh H, Shannon SE. Three approaches to qualitative content analysis. *Qual Health Res* 2005 Nov;15(9):1277-1288. [doi: [10.1177/1049732305276687](https://doi.org/10.1177/1049732305276687)] [Medline: [16204405](https://pubmed.ncbi.nlm.nih.gov/16204405/)]
20. Fairburn CG, Cooper Z. The eating disorder examination. In: Fairburn CG, Wilson G, editors. *Binge eating: nature, assessment, and treatment*. New York: Guilford Press; 1993:317-360.
21. Freedman MJ, Lester KM, McNamara C, Milby JB, Schumacher JE. Cell phones for ecological momentary assessment with cocaine-addicted homeless patients in treatment. *J Subst Abuse Treat* 2006 Mar;30(2):105-111. [doi: [10.1016/j.jsat.2005.10.005](https://doi.org/10.1016/j.jsat.2005.10.005)] [Medline: [16490673](https://pubmed.ncbi.nlm.nih.gov/16490673/)]

## Abbreviations

**DSM-IV:** Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition

**EMA:** ecological momentary assessment

**IRB:** institutional review board

**MA:** Mexican American

**NVS:** Newest Vital Sign

**SAHL:** Short Assessment of Health Literacy

**SES:** socioeconomic status

**UEWCB:** unhealthy eating and weight control behavior

*Edited by G Eysenbach; submitted 08.01.16; peer-reviewed by J Beaulac, J Jones, S Iribarren; comments to author 04.02.16; revised version received 17.03.16; accepted 18.03.16; published 14.07.16.*

*Please cite as:*

*Connelly K, Stein KF, Chaudry B, Trabold N*

*Development of an Ecological Momentary Assessment Mobile App for a Low-Literacy, Mexican American Population to Collect Disordered Eating Behaviors*

*JMIR Public Health Surveill* 2016;2(2):e31

URL: <http://publichealth.jmir.org/2016/2/e31/>

doi: [10.2196/publichealth.5511](https://doi.org/10.2196/publichealth.5511)

PMID: [27418020](https://pubmed.ncbi.nlm.nih.gov/27418020/)

©Kay Connelly, Karen F. Stein, Beenish Chaudry, Nicole Trabold. Originally published in *JMIR Public Health and Surveillance* (<http://publichealth.jmir.org>), 14.07.2016. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in *JMIR Public Health and Surveillance*, is properly cited. The complete bibliographic information, a link to the original publication on <http://publichealth.jmir.org>, as well as this copyright and license information must be included.

Original Paper

# Qualitative Analysis of Cognitive Interviews With School Children: A Web-Based Food Intake Questionnaire

Vanessa Fernandes Davies<sup>1\*</sup>, PhD; Emil Kupek<sup>1</sup>, PhD; Patricia Faria Di Pietro<sup>2\*</sup>, PhD; Maria Alice Altenburg de Assis<sup>2\*</sup>, PhD; Francilene GK Vieira<sup>2\*</sup>, PhD; Clarice Perucchi<sup>2\*</sup>, RD; Rafaella Mafra<sup>2\*</sup>, RD; Debbe Thompson<sup>3\*</sup>, PhD; Thomas Baranowski<sup>3\*</sup>, PhD

<sup>1</sup>Public Health, Santa Catarina Federal University, Florianopolis, Brazil

<sup>2</sup>Nutrition Department, Santa Catarina Federal University, Florianopolis, Brazil

<sup>3</sup>Department of Pediatrics, Baylor College of Medicine, Houston, TX, United States

\*these authors contributed equally

**Corresponding Author:**

Vanessa Fernandes Davies, PhD

Public Health

Santa Catarina Federal University

R. Delfino Conti, S/N - Trindade

Florianopolis, 88036-020

Brazil

Phone: 55 (48) 3721 9394

Fax: 55 (48) 3721 9394

Email: [va.davies@hotmail.co.uk](mailto:va.davies@hotmail.co.uk)

## Abstract

**Background:** The use of computers to administer dietary assessment questionnaires has shown potential, particularly due to the variety of interactive features that can attract and sustain children's attention. Cognitive interviews can help researchers to gain insights into how children understand and elaborate their response processes in this type of questionnaire.

**Objective:** To present the cognitive interview results of children who answered the WebCAAFE, a Web-based questionnaire, to obtain an in-depth understanding of children's response processes.

**Methods:** Cognitive interviews were conducted with children (using a pretested interview script). Analyses were carried out using thematic analysis within a grounded theory framework of inductive coding.

**Results:** A total of 40 children participated in the study, and 4 themes were identified: (1) the meaning of words, (2) understanding instructions, (3) ways to resolve possible problems, and (4) suggestions for improving the questionnaire. Most children understood questions that assessed nutritional intake over the past 24 hours, although the structure of the questionnaire designed to facilitate recall of dietary intake was not always fully understood. Younger children (7 and 8 years old) had more difficulty relating the food images to mixed dishes and foods eaten with bread (eg, jam, cheese). Children were able to provide suggestions for improving future versions of the questionnaire.

**Conclusions:** More attention should be paid to children aged 8 years or below, as they had the greatest difficulty completing the WebCAAFE.

(*JMIR Public Health Surveill* 2016;2(2):e167) doi:[10.2196/publichealth.5024](https://doi.org/10.2196/publichealth.5024)

**KEYWORDS**

dietary assessment; children; computer; questionnaire

## Introduction

Dietary assessment among children represents a challenge to nutritional epidemiology due to systematic errors (ie, gender, weight, irregular food patterns, social desirability) [1-4]. In

addition, errors related to children's cognitive development can impact their memory of foods they have eaten and their ability to differentiate between dishes and estimate food portions [5,6].

Web-based questionnaires for assessing children's food intake have been developed as an alternative to traditional dietary recall

methods [7-12]. The advantages of Web-based questionnaires include the following: the extensive interactive resources that can be used to increase children's motivation to complete the questionnaire, as well as to hold their attention; the standardization of questions, allowing for better quality control; the possibility of self-reporting by children, which reduces the cost of selecting and training interviewers; and the possibility of obtaining quick results [13,14]. A common technique in Web-based questionnaires is to use a character (an avatar) to interact with, guide, or assist children with their recall [15].

In Brazil, the System for Monitoring Food Intake and Physical Activity of Schoolchildren—the WebCAAFE—was developed [16] as a Web-based questionnaire that used an avatar in the form of a robot to assist children with their recall. Its purpose was to enable schoolchildren to self-report their food consumption and physical activity (24-hour recall) within their school environment. Formative and usability studies were conducted [17-19], and the WebCAAFE section related to food intake was validated [8]. Direct observation of school lunch as a reference method showed that the WebCAAFE presented similar validity to other Web-based food intake questionnaires [10,20] (43% matches, 29% intrusions, and 28% omissions), and there was greater disagreement in the responses of younger children (<8 years) compared with older children [8].

Since children receive instructions from the avatar to self-report food intake, ideally having no contact with the researcher, 3 important points should be addressed. First, the words and instructions should be clearly understood by those being interviewed. Technical terms that are not part of the children's daily life can be problematic. Second, questions should be designed taking into account the respondents' cognitive skills and daily activities. Finally, the answer options must be easily understood by the children [21,22]. The WebCAAFE validation studies raised some important issues about the children's ability to understand and solve problems when answering the questionnaire. Therefore, to respond to the issues raised and improve children's performance in the WebCAAFE questionnaire in future studies, cognitive interviews were conducted as a follow-up to the validation studies. The researchers also expected to gain some insights into how Brazilian children answered food questionnaires. The majority of research using qualitative methods to improve food intake questionnaires has been conducted in different sociocultural contexts [11,23-26], and cognitive interviews could fill an important gap in Brazil and similar contexts.

Therefore, this study presents the cognitive interview results of children who answered the WebCAAFE, to obtain an in-depth understanding of children's response processes.

## Methods

### Participants

An intentional sample [27] of children (based on age and gender) from second to fifth grade (approximately 7-11 years old) at a public primary school in Florianópolis, Brazil, was invited to

participate in the cognitive interviews. The ages of Brazilian schoolchildren in specific grades were usually as follows: second grade: 7-8 years; third grade: 8-9 years; fourth grade: 9-10 years; fifth grade: 10-11 years. Ten children participated from each grade (total: 40 children, 21 boys). The choice of school where the study would be conducted was based on the existence of a suitable computer room and Internet connection for the interviews to be carried out. The number of children chosen to be interviewed was determined by the principle of saturation of the collected data (ie, saturation of information was the criterion used to determine the end of the qualitative data collection) [27]. The teachers were responsible for choosing the participants based on the probability that their parents would give their consent for their children to participate. A consent form was sent to the parents or guardians to authorize the participation of their children in the study. Children whose parents did not provide consent (n=6) were replaced by others. At the time of the interview, the children's assent was also required for them to participate. The project was approved by the Ethics Committee on Human Research at the Federal University of Santa Catarina (UFSC) (Protocol 2250/11).

### The WebCAAFE Questionnaire

The WebCAAFE questionnaire was a single-day recall procedure from the previous day divided into 3 sections: (1) registration form, (2) diet, and (3) physical activity [19]. The registration section referred to information about respondents, such as their name, mother's name, sex, weight, height, age, date of birth, and study period. The physical activity section was divided into the 3 parts of the day (morning, afternoon, and evening). Children could choose sedentary and physical activity icons from 32 options.

The food intake section, the objective of this study, was divided into 6 eating occasions (breakfast, morning snack, lunch, afternoon snack, dinner, and evening snack). For each meal, 32 pictures of foods or beverages were displayed on the computer screen so that the child could make their selections (Figure 1). The foods groups, including healthy and unhealthy items, were chosen taking into account the food patterns of children in this age group (reported by schoolchildren in the 7-day food diaries), foods offered in school meals, and foods recommended in the Brazilian food guidelines [19]. A robot avatar guided the child during the interview, explaining the concept of each meal and the time of day at which it was normally consumed, and reinforced the importance of only reporting food intake from the day before (Figure 2). On completing the answers, the child was offered the chance to review the selected foods and make any necessary changes (Figure 3). After completing the 6 eating events, questions about school meals were asked on different screens: (1) "Did you have a school meal yesterday?" (yes/no); (2) if yes, "Which of the foods you selected did you eat during the school meal?" The WebCAAFE system has a database of 350 icons of foods and beverages, as well as 50 icons of different physical activities or sedentary behaviors. This database allows for the selection of 32 icons of foods/beverages and 32 icons of physical activities or sedentary behaviors.

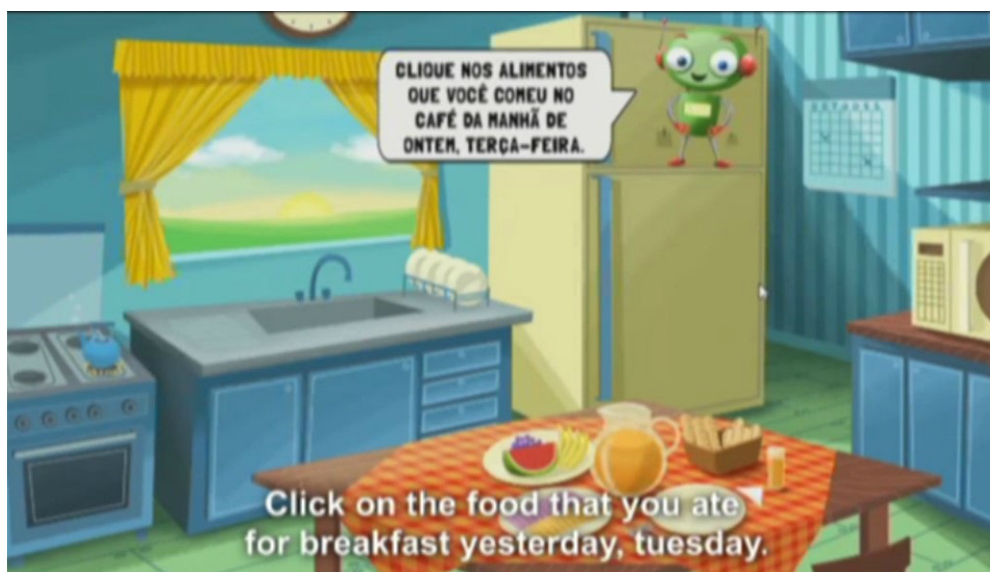


Figure 1. Screenshot of the WebCAAFE questionnaire food options from which the children chose what they consumed the previous day.

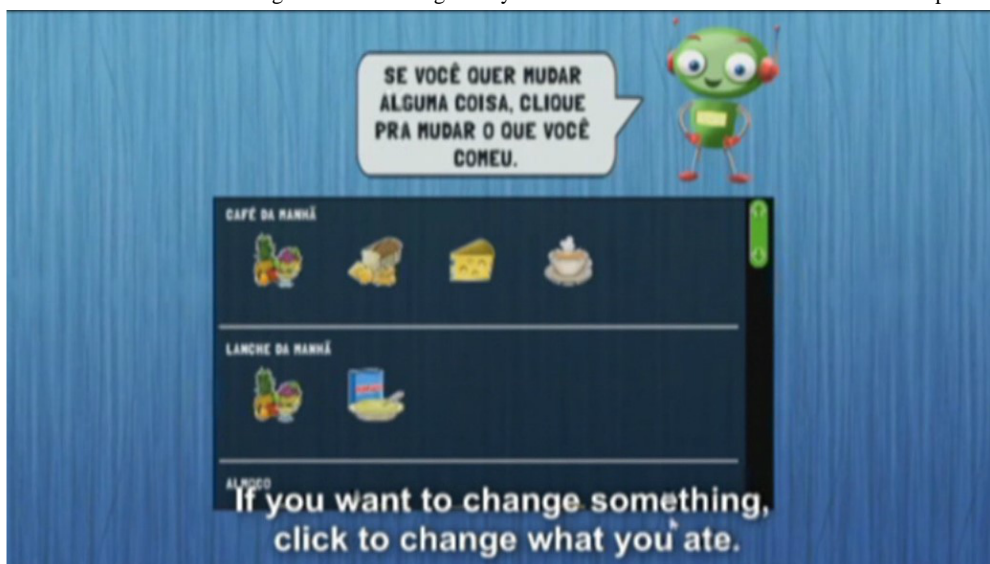




**Figure 2.** Screenshot of the WebCAAFE showing the instructions given by the avatar for the children to report their food consumption in the previous 24 hours.



**Figure 3.** Screenshot of the WebCAAFE showing the instructions given by the avatar for the children to review the foods previously selected.



### Data Collection

The cognitive interviews were conducted individually between August and September 2014 in the computer room at the participants' school. The average session took 60 minutes. The interviews were conducted by one of the authors (VFD) and undergraduate nutrition students ( $n=6$ ) who had been trained in qualitative techniques and cognitive interviews (4 training sessions, mean duration of sessions: 90 minutes). Before collecting data, interviewers practiced the cognitive interviews with 6 children of similar age to those in the study. Cognitive

interviews were conducted at the same time as children were answering the WebCAAFE Web-based questionnaire. The interview script is presented in the [Table 1](#). To ensure consistency in the data collection, the interview script was developed according to the following structure: main question; question to expand and clarify the child's response; and follow-up question (to link the children's answers to possible suggestions for improving the WebCAAFE). The script was piloted in a test with 10 children from second grade (5 boys and 5 girls) to rephrase or refine the questions in the cognitive interview.

**Table 1.** Examples of questions on the WebCAAFE questionnaire and guides for conducting main questions, probes, and follow-up questions during cognitive interviews sessions with schoolchildren.

WebCAAFE question asked by the Avatar	Aim of the question	Main question	Prompts	Follow-up question
Breakfast is the first meal of the day	To know whether the child understands the word meal and know the meals of the day	Can you tell me in your own words what the Avatar is saying to you ?	<p>What does the word meal mean to you? Can you give me one example?</p> <p>Is breakfast an example of a meal? Can you give other examples of meals?</p> <p>- For example, you said that breakfast is a meal. When do you think you have it (that meal)? Where do you have it?</p>	<p>Do you think we should change any words so the children can understand better?</p> <p>-For example, you said that you didn't understand breakfast. How could you make it better?</p>
Breakfast is the first meal of the day after we wake up. Click on foods that you ate for breakfast yesterday. Remember, if you did not eat anything, click on the 'nothing' button	To know if the children understand instructions given by the Avatar. In other words, they should know that they need to answer about foods/drinks they consumed for breakfast yesterday; if they did not eat anything or if they do not remember, the correct answer should be 'nothing'*.	Can you tell me in your own words what the Avatar is saying to you?	<p>How do you choose something that you ate?</p> <p>If you didn't eat anything, but drank juice, what should you do?</p> <p>Is the Avatar asking about today or another day?</p>	<p>- Which other ways do you suggest?</p> <p>- Which word or words do you suggest to improve this question?</p>
Breakfast is the first meal of the day after we wake up. Click on foods that you ate for breakfast yesterday. Remember, if you did not eat anything, click on the 'nothing' button	To understand how children's response processes work	Can you tell me why you chose those foods/drinks?	<p>-Did you find everything that you were looking for?</p> <p>-Can you give me an example of how you would answer if you had a cheese sandwich/tuna pie/pasta with bolognese sauce/a sandwich with chicken paté/an omelet/ bread and jam/vegetable pie</p> <p>-Let's say that you ate something, but the image of that food is not on screen. What would you do?</p> <p>-If you can't remember what you ate yesterday, what should you do?</p>	
Now, let's talk about the morning snack**. This is the meal that you eat after breakfast and before lunch. This would be the meal that you normally have at school. Click on foods that you ate for the morning snack yesterday. Remember, if you did not eat anything, click on the 'nothing' button	To understand how children answer questions about foods that they consumed at school	Can you tell me why you chose those foods/drinks?	<p>-Let's say that you brought food from home. How would you answer that question?</p> <p>- How would you answer this question if you did not come to school yesterday?</p> <p>- How would you answer if you came to school yesterday, but did not eat anything?</p>	
If you want to change anything, click on the button to change what you ate/drank. If you don't want to change anything, click on 'continue'	To verify if the children understand instructions. In other words, they should know that the foods presented on the screen are the ones that they chose previously, and at this time they should decide if they want to change any of the answers they gave related to food consumption	Can you tell me in your own words what the Avatar is saying to you ?	<p>Do you recognize these foods/drinks?</p> <p>-Take a look at the screen. Are these the foods and drinks that you chose earlier? (ask this question only if the children do not recognize the foods on the screen)</p> <p>- What do you think you should do now?</p> <p>- What do you think you should do if you want to change something?</p> <p>- What should you do if you don't want to change anything?</p> <p>- What do you think you should do if you just remembered a food or drink that you had yesterday, but that you forgot to choose?</p>	

\* This question should be asked only after the child chooses the foods consumed, or if the child decided to click on the 'nothing' button.

\*\* If the child goes to school in the afternoon, this question should be asked about the afternoon snack.

## Data Analyses

All the interviews were audio recorded and transcribed verbatim by the interviewers. The transcripts were compared with the respective audio recordings by 2 researchers (VFD and MRI) to verify accuracy. The same researchers coded the interviews individually (manually), using a thematic matrix technique, within a grounded theory framework of inductive coding [28]. The thematic matrix method included the following steps: rereading the transcripts to identify themes for the organization of data; indexing the themes within the transcripts; and removing data from the transcripts and transferring them to a matrix theme (each theme relating to quotes from the participants). The matrix included the main ideas and perceptions of the individual related to the theme. On completing the coding, analyses were compared, discrepancies were discussed, and a consensus was reached.

## Results

### Principal Findings

A total of 40 children participated in the interviews (mean age: second grade, 7.8 years; third grade, 8.9 years; fourth grade, 9.4 years; fifth grade, 10.8 years; 21 boys). Four themes were identified: (1) the meaning of words, (2) understanding instructions, (3) ways to resolve possible problems, and (4) suggestions for improving the questionnaire. Throughout this study, the children's comments were identified by the letter C (short for the word "child") and their school grade to preserve the participants' confidentiality.

#### Theme 1: The Meaning of Words

The words specifically investigated were "meal" and "school meal." The aim of investigating the term "meal" was to assess the child's recognition of eating occasions (meals or snacks) during the day, the time at which they were consumed, and the types of food consumed at each meal. The term "school meal" was tested to ascertain whether the children associated these words with meals consumed at school, as the WebCAAFE also sought to record information about food consumed at school.

Most children did not know the meaning of the word "meal." Regardless of age, when asked about this word children cited examples of foods or said the meaning was the same as "food." When subsequent questions (probes) were asked to expand the children's answers (the interviewer asked specifically about each meal, for example: "Do you know what breakfast is?"), children cited examples of foods normally consumed in those meals (eg, coffee and bread for breakfast, rice and beans for lunch). Children also correctly identified the time of day and where breakfast, lunch, and dinner were eaten. Eating between meals, such as morning, afternoon, or evening snacks, was recognized only when the snack was consumed at school.

*Is there an evening snack as well? I didn't even know we had stuff like that. [C, second grade]*

*I've never used the words evening snack...afternoon snack...morning snack...I don't say anything like that...I just say dinner and lunch...and breakfast...when I wake up early. [C, fourth grade]*

*Children aren't used to having some meals and they would get confused when the Avatar asks about afternoon snack, morning snack and evening snack [C, fourth grade]*

The word "school meal" was understood by the children to mean all food prepared by the kitchen staff.

#### Theme 2: Understanding Instructions

This category included children's responses regarding their comprehension of the instructions given by the avatar to answer questions about their food consumption and review the foods that they had previously chosen.

In general, children understood the directions and what they were being asked to do. Few children did not know the period (yesterday) about which they should report.

The children who did not usually eat a morning or evening snack were often confused by the question about food consumption for these eating occasions.

*"I never knew there was one..." (Comment from a child when asked by the avatar to report food intake in the morning snack) [C, third grade]*

The snack consumed at school was an important reference for the child to know which foods they had consumed at what time. When asked about this meal, the children seemed to understand that they should record any food consumed at school, whether in school meals, bought in the school canteen, or brought from home. The most frequent doubts were about situations in which children were not at school at the time when the school snack was consumed and did not know whether they should still answer what they had consumed at home or elsewhere.

One instruction given by the avatar that appeared to be understood by most children was when they were asked to check their answers and change anything they thought necessary.

#### Theme 3: Ways to Resolve Possible Problems

This category included the ways in which the children answered the questionnaire when they did not remember what they had eaten the previous day or when they could not find the food images that represented their food intake. This category also recorded how the children answered the questions relating to eating occasions that were not part of their daily routine or that they did not know, such as the morning and evening snacks.

Seven-year-old children who did not remember what they had eaten the day before selected foods at random:

*"I'll pretend I ate it." (Child clicked randomly on various foods) [C, third grade]*

Older children (9 years old and above) who did not remember what they had consumed selected the "nothing" button. However, there were some situations in which this age group did not remember what they had consumed and they considered clicking on foods that were usually eaten.

*I don't remember what I ate, so I thought about choosing soup, which is what I have when it's cold...but then I thought again and decided to tell the*



*truth, which is: I don't remember what I had.* [C, fourth grade]

*Younger children, when they could not find the picture of the food they were looking for, chose other foods to compensate. "Then I'll put juice, it's easier. Juice is the same as water, it's made of water."* [C, second grade]

For the group of children aged 7 and 8 years, another possibility to compensate for the lack of an image would be to ask for an adult's help. In some interviews, children were heard to comment that they would ask the teacher when they did not know what to do or they could not find the food they had consumed. During the interview, some food images in the WebCAAFE were not recognized by the children, such as chicken nuggets, the pasta group, cheese bread, manioc flour, porridge, and mussels.

To understand children's response processes, interviewers asked the children to locate foods on the WebCAAFE screen usually found on the school menu. The foods investigated were cheese sandwich, bread with jam, bread with chicken pâté, vegetable pie, tuna pie, and pasta with minced beef. To locate these foods, the children would ideally make an association with the individual components of the dish (eg, pasta with minced beef—the child would select the pasta and the meat separately, or would select the main component (eg, bread with jam, the child would select at least the bread, as there was no image of jam).

The children's answers to this question varied by age. Among younger children (7 or 8 years old), some chose food at random that was in no way similar to the investigated food. Others associated foods that were markers of healthy eating, such as bread and cheese with the group of pizzas and hamburgers, or pasta and minced beef with noodles and sausages. A few others reported that the food image could not be found in the WebCAAFE. In addition, many children identified the main food on the screen, such as pasta or bread.

*"If I'd eaten bread and cheese, I'd just answer bread, and for pasta with minced beef, I'd only answer pasta."* [C, third grade]

It seemed that children aged 9 years or above tried to choose the most appropriate image even if in some cases, the nutritional content of the foods was not fully comparable. For example, bread and fish were chosen to represent tuna pie; bread and sweets represented bread with jam; bread and fruit represented bread with jam; the vegetables group represented vegetable pie; pasta and vegetables represented vegetable pie. Children in the same age group wanted to provide details of preparation, even adding specific ingredients (eg, onions, sour cream).

*"It's the way my grandma makes omelet, she makes it with onion and carrot" (justifying the answer given for what she had eaten: omelet).* [C, fifth grade]

With regard to the morning and evening snacks, 2 response patterns were observed: (1) reporting consumption of a particular food even if they had previously said during the interview that they did not know what these snacks were or did not eat at those times and (2) confusion among the children relating the morning

snack and the evening snack to breakfast or with the dessert at dinner.

#### **Theme 4: Suggestions for Improving the Questionnaire**

This category included the suggestions made by the children for improving the WebCAAFE. Although the WebCAAFE showed the name of the food when the child passed the cursor over the image on the screen, this detail was not always noticed by the children. For situations in which the food images were not recognized, some children suggested subtitles:

*"You could add something below the images to describe what the foods are, for example, to write coffee below the coffee image, soup below the soup image..."* [C, fifth grade]

*Should there not be an image of a food consumed, children suggested writing the name of the food in the space provided: "It would also be nice to write...I would like to write...the foods, if there wasn't a food you could write it down."* [C, fourth grade]

Another suggestion was the inclusion of more food illustrations:

*At least put water. You can put more things, if there aren't many things we won't know what we ate or what we didn't eat. Then, when we remember what we ate, we need to have more choices, because only these things, we hardly eat any of them.* [C, second grade]

Regarding the understanding of instructions, one child suggested offering the option of listening to the avatar's instructions again:

*"I think you could do that...the child didn't understand...goes there and clicks on the little robot and he appears again, speaking so the child can understand better."* [C, third grade]

## **Discussion**

### **Principal Findings**

The main results from the cognitive interviews revealed important insights into how children understood and prepared answers to the avatar's questions, as well as suggestions for improving the questionnaire. The eating occasions which were most recognized by the children were breakfast, the snack eaten at school, lunch, and dinner. Instructions on how to report food intake were also understood by the majority of children, as well as the instructions to review their selected foods at the end of the questionnaire. Younger children were less able to solve problems and were less consistent in their response processes compared with older children. These results answered questions raised by the WebCAAFE researchers during the validation study, as well as reflected on the possible ways of dealing with children's difficulties in future applications of the questionnaire.

It has been recommended that Web-based questionnaires to assess children's food intake follow a guided meal-based approach to help children recall information [11,12,29]. Our interviews, however, showed that for the sample of children participating in this study, eating occasions such as morning and evening snacks were either not part of their usual routine or not known by the children, and may therefore be confused

with other meals. This problem was also identified in a formative study with nutritionists for the development of the WebCAAFE [17]. Children who did not consume snacks at these times were induced to respond that they did consume them, thus increasing errors. These results indicated that children needed a more detailed explanation of the types of meals before completing the WebCAAFE. When the questionnaire was administered at schools, researchers followed a written protocol to explain how to answer the questionnaire in the classroom with children, for which a banner (140 × 105 cm) was used to illustrate the foods and meals included in the WebCAAFE questions. A possible solution would be to explain more clearly the precise meaning of these meals and make it clear to the children that if the meal was not eaten, they should not choose any food.

In Brazil, there is the school feeding program as well as laws prohibiting the sale of processed foods or soft drinks in schools [30]. The results from the cognitive interviews showed that the school meal was an important reference for children, both to help them understand what time of day the avatar was asking about and to remember the foods they had consumed. This is considered a positive result as it shows the potential of the WebCAAFE for assessing food consumed in the school environment, thereby serving as an important tool for evaluating nutrition policies.

Similar to the WebCAAFE validation study, children aged 9 years and above performed better than younger children in this study (4). Children aged 7 and 8 years had more difficulty when searching for images of food that reflected their consumption more literally. This information was in accordance with other studies that experienced the same problem with regard to children's reporting of mixed dishes, different types of preparation (eg, boiled egg or fried egg), or foods that can be added to other foods (such as jam, cheese, butter, or mayonnaise) [6-10,31-33,34]. For younger children answering the WebCAAFE, the presence of an adult (teacher or researcher) to monitor and solve problems during the administration of the questionnaire could minimize errors. Since some children commented that they would ask the teacher when they did not know what to do or could not find the food consumed, adults could use the school menu as a reference to guide the children to answer the questions accurately. For example, if children had consumed vegetable pie, they could be advised to click on the vegetables group as well as on the breads and crackers group. Standardized guidelines could also be developed to specify procedure in the absence of images (ie, when the child consumed something for which there was no image in the WebCAAFE, they should be advised to click on the "nothing" button). Adult assistance for children younger than 8 years reporting food intake was also recommended in other studies [24,25].

A good understanding of the avatar's instructions was a positive result, confirming the importance of the research conducted during the development of the WebCAAFE, such as the focus group discussions with nutritionists and the questionnaire usability studies [17,19]. However, should there be future studies with the WebCAAFE in other regions of Brazil, it is important to reevaluate this issue, as there are major socioeconomic differences in the country. It may be necessary to train children

before answering the questionnaire so they are better able to answer the questions and solve problems that may arise.

Changes in response to children's suggestions could contribute to further improvements, for example, the use of more familiar words for children (eg, "school snack" instead of "school meal"), the use of more visible subtitles to identify the food images, more common food options (eg, water), and the inclusion of a space to write the foods consumed. Other Web-based food intake assessment questionnaires have used a write-in feature [10,11]. It should be emphasized that the WebCAAFE system has a database of 350 foods from which 32 food icons can be selected to be displayed on the screen, so the food items can be replaced by others depending on the research and objectives. The findings of this study, as well as discussions with specialists in educational psychology and software development, may help to improve the next version of the questionnaire with explanations that take into account the limitations of children's vocabulary and add the option for children to write a food if it is not included on the screen.

The strengths of this study included conducting a pilot study to modify and refine the interview questions; the use of an interview script to ensure consistency of the collected data; carrying out research in a school similar to those where the questionnaire would be administered; and extensive training of the interviewers. A limitation of this study was conducting the study in one area, which did not allow for generalizations to be made about the children's cognitive performance when answering questionnaires in other sociocultural contexts. Regional variations in the school menus in Brazil may also lead to other difficulties for children to answer the same questionnaire.

Future studies should evaluate the addition of an adult as a facilitator when children aged 7-8 years are completing the questionnaire, guiding them through the WebCAAFE screens one by one.

## Conclusions

This study obtained similar results to other studies with Web-based tools for the assessment of food consumption. However, this did not alter the relevance of this study, as the cognitive interviews provided not only important insights for improving future versions of the WebCAAFE questionnaire but also general information about Brazilian children answering Web-based food questionnaires, contributing to the research on this issue in a country with little experience of this type of technological innovation.

The questionnaire structure (6 eating occasions per day) was not understood by many children, as some meals were not part of their normal routines or were unknown to them. Choosing images of mixed dishes and foods eaten with bread (eg, jam) were complex tasks that demanded cognitive skills that younger children might not have fully developed. Finally, as indicated in the WebCAAFE validation study, more attention should be paid to children aged 8 years and below, as they had the greatest difficulty completing the WebCAAFE. The assistance of an adult when younger children are completing the questionnaire



could be an alternative way of improving the accuracy of the answers.

## Acknowledgments

This research was funded by a grant from the Brazilian Ministry of Health (Departamento de Ciência, Tecnologia e Insumos Estratégicos—DECIT), the Brazilian Ministry of Education (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior—CAPES), and the Brazilian Ministry of Science, Technology and Innovation (Conselho Nacional de Desenvolvimento Científico e Tecnológico—CNPq). The authors gratefully thank the Education Secretary of the Prefeitura Municipal de Florianópolis, the staff of the schools, and children and parents for their participation in the study.

## Conflicts of Interest

None declared.

## References

1. Murakami K, Livingstone MB. Prevalence and characteristics of misreporting of energy intake in US children and adolescents: National Health and Nutrition Examination Survey (NHANES) 2003-2012. *Br J Nutr* 2016 Jan 28;115(2):294-304. [doi: [10.1017/S0007114515004304](https://doi.org/10.1017/S0007114515004304)] [Medline: [26525591](https://pubmed.ncbi.nlm.nih.gov/26525591/)]
2. Lioret S, Touvier M, Balin M, Huybrechts I, Dubuisson C, Dufour A, et al. Characteristics of energy under-reporting in children and adolescents. *Br J Nutr* 2011 Jun;105(11):1671-1680. [doi: [10.1017/S0007114510005465](https://doi.org/10.1017/S0007114510005465)] [Medline: [21262062](https://pubmed.ncbi.nlm.nih.gov/21262062/)]
3. Collins CE, Watson J, Burrows T. Measuring dietary intake in children and adolescents in the context of overweight and obesity. *Int J Obes (Lond)* 2010 Jul;34(7):1103-1115. [doi: [10.1038/ijo.2009.241](https://doi.org/10.1038/ijo.2009.241)] [Medline: [19935750](https://pubmed.ncbi.nlm.nih.gov/19935750/)]
4. de Assis MA, Kupek E, Guimarães D, Calvo MC, de Andrade DF, Bellisle F. Test-retest reliability and external validity of the previous day food questionnaire for 7-10-year-old school children. *Appetite* 2008 Jul;51(1):187-193. [doi: [10.1016/j.appet.2008.02.014](https://doi.org/10.1016/j.appet.2008.02.014)] [Medline: [18375017](https://pubmed.ncbi.nlm.nih.gov/18375017/)]
5. Baxter SD. Cognitive processes in children's dietary recalls: insight from methodological studies. *Eur J Clin Nutr* 2009 Feb;63 Suppl 1:S19-S32 [FREE Full text] [doi: [10.1038/ejcn.2008.61](https://doi.org/10.1038/ejcn.2008.61)] [Medline: [19190640](https://pubmed.ncbi.nlm.nih.gov/19190640/)]
6. Livingstone MB, Robson PJ. Measurement of dietary intake in children. *Proc Nutr Soc* 2000 May;59(2):279-293. [Medline: [10946797](https://pubmed.ncbi.nlm.nih.gov/10946797/)]
7. Medin AC, Astrup H, Kåsin BM, Andersen LF. Evaluation of a Web-Based Food Record for Children Using Direct Unobtrusive Lunch Observations: A Validation Study. *J Med Internet Res* 2015 Dec 07;17(12):e273 [FREE Full text] [doi: [10.2196/jmir.5031](https://doi.org/10.2196/jmir.5031)] [Medline: [26680744](https://pubmed.ncbi.nlm.nih.gov/26680744/)]
8. Davies VF, Kupek E, de Assis MA, Natal S, Di Pietro PF, Baranowski T. Validation of a web-based questionnaire to assess the dietary intake of Brazilian children aged 7-10 years. *J Hum Nutr Diet* 2015 Jan;28(Suppl 1):93-102. [doi: [10.1111/jhn.12262](https://doi.org/10.1111/jhn.12262)] [Medline: [25139011](https://pubmed.ncbi.nlm.nih.gov/25139011/)]
9. Carvalho MA, Baranowski T, Foster E, Santos O, Cardoso B, Rito A, et al. Validation of the Portuguese self-administered computerised 24-hour dietary recall among second-, third- and fourth-grade children. *J Hum Nutr Diet* 2015 Dec;28(6):666-674. [doi: [10.1111/jhn.12280](https://doi.org/10.1111/jhn.12280)] [Medline: [25420921](https://pubmed.ncbi.nlm.nih.gov/25420921/)]
10. Baranowski T, Islam N, Douglass D, Dadabhoj H, Beltran A, Baranowski J, et al. Food Intake Recording Software System, version 4 (FIRSS4): a self-completed 24-h dietary recall for children. *J Hum Nutr Diet* 2014 Jan;27(Suppl 1):66-71 [FREE Full text] [doi: [10.1111/j.1365-277X.2012.01251.x](https://doi.org/10.1111/j.1365-277X.2012.01251.x)] [Medline: [22616645](https://pubmed.ncbi.nlm.nih.gov/22616645/)]
11. Billoft-Jensen A, Trolle E, Christensen T, Islam N, Andersen LF, Egenfeldt-Nielsen S, et al. WebDASC: a web-based dietary assessment software for 8-11-year-old Danish children. *J Hum Nutr Diet* 2014 Jan;27(Suppl 1):43-53. [doi: [10.1111/j.1365-277X.2012.01257.x](https://doi.org/10.1111/j.1365-277X.2012.01257.x)] [Medline: [22594587](https://pubmed.ncbi.nlm.nih.gov/22594587/)]
12. Moore HJ, Ells LJ, McLure SA, Crooks S, Cumbor D, Summerbell CD, et al. The development and evaluation of a novel computer program to assess previous-day dietary and physical activity behaviours in school children: the Synchronised Nutrition and Activity Program (SNAP). *Br J Nutr* 2008 Jun;99(6):1266-1274. [doi: [10.1017/S0007114507862428](https://doi.org/10.1017/S0007114507862428)] [Medline: [18042307](https://pubmed.ncbi.nlm.nih.gov/18042307/)]
13. Touvier M, Kesse-Guyot E, Méjean C, Pollet C, Malon A, Castetbon K, et al. Comparison between an interactive web-based self-administered 24 h dietary record and an interview by a dietitian for large-scale epidemiological studies. *Br J Nutr* 2011 Apr;105(7):1055-1064. [doi: [10.1017/S0007114510004617](https://doi.org/10.1017/S0007114510004617)] [Medline: [21080983](https://pubmed.ncbi.nlm.nih.gov/21080983/)]
14. Ngo J, Engelen A, Molag M, Roesle J, García-Segovia P, Serra-Majem L. A review of the use of information and communication technologies for dietary assessment. *Br J Nutr* 2009 Jul;101(Suppl 2):S102-S112. [doi: [10.1017/S0007114509990638](https://doi.org/10.1017/S0007114509990638)] [Medline: [19594959](https://pubmed.ncbi.nlm.nih.gov/19594959/)]
15. Lu AS, Baranowski J, Islam N, Baranowski T. How to engage children in self-administered dietary assessment programmes. *J Hum Nutr Diet* 2014 Jan;27(Suppl 1):5-9 [FREE Full text] [doi: [10.1111/j.1365-277X.2012.01258.x](https://doi.org/10.1111/j.1365-277X.2012.01258.x)] [Medline: [22594618](https://pubmed.ncbi.nlm.nih.gov/22594618/)]
16. CAAFE. Sistema de monitoramento do consumo alimentar e atividade física de escolares. URL: <http://caafe.ufsc.br/> [accessed 2016-11-02] [WebCite Cache ID 6ljGDwvWH]

17. Davies VF, Kupek E, de Assis MA, Engel R, da Costa FF, Di Pietro PF, et al. Qualitative analysis of the contributions of nutritionists to the development of an online instrument for monitoring the food intake of schoolchildren. *J Hum Nutr Diet* 2015 Jan;28(Suppl 1):65-72. [doi: [10.1111/jhn.12209](https://doi.org/10.1111/jhn.12209)] [Medline: [24480047](https://pubmed.ncbi.nlm.nih.gov/24480047/)]
18. da Costa FF, Davies VF, Schmoelz CP, Kuntz MGF, de Assis MAA. Medida da atividade física de crianças: o que professores de educação física têm a nos dizer? *Rev Bras Ativ Fis e Saude* 2012;17(4):286-292.
19. da Costa FF, Schmoelz CP, Davies VF, Di Pietro PF, Kupek E, de Assis MA. Assessment of diet and physical activity of brazilian schoolchildren: usability testing of a web-based questionnaire. *JMIR Res Protoc* 2013;2(2):e31 [[FREE Full text](#)] [doi: [10.2196/resprot.2646](https://doi.org/10.2196/resprot.2646)] [Medline: [23958804](https://pubmed.ncbi.nlm.nih.gov/23958804/)]
20. Baranowski T, Islam N, Baranowski J, Cullen KW, Myres D, Marsh T, et al. The food intake recording software system is valid among fourth-grade children. *J Am Diet Assoc* 2002 Mar;102(3):380-385. [Medline: [11902371](https://pubmed.ncbi.nlm.nih.gov/11902371/)]
21. Collins D. Pretesting survey instruments: an overview of cognitive methods. *Qual Life Res* 2003 May;12(3):229-238. [Medline: [12769135](https://pubmed.ncbi.nlm.nih.gov/12769135/)]
22. Carlton J. Developing the draft descriptive system for the child amblyopia treatment questionnaire (CAT-Qol): a mixed methods study. *Health Qual Life Outcomes* 2013;11:174 [[FREE Full text](#)] [doi: [10.1186/1477-7525-11-174](https://doi.org/10.1186/1477-7525-11-174)] [Medline: [24148800](https://pubmed.ncbi.nlm.nih.gov/24148800/)]
23. Hanning RM, Royall D, Toews JE, Blashill L, Wegener J, Driezen P. Web-based Food Behaviour Questionnaire: validation with grades six to eight students. *Can J Diet Pract Res* 2009;70(4):172-178. [doi: [10.3148/70.4.2009.172](https://doi.org/10.3148/70.4.2009.172)] [Medline: [19958572](https://pubmed.ncbi.nlm.nih.gov/19958572/)]
24. Vereecken C, Covents M, Maes L, Moysen T. Formative evaluation of the dietary assessment component of Children's and Adolescents' Nutrition Assessment and Advice on the Web (CANAA-W). *J Hum Nutr Diet* 2014 Jan;27(Suppl 1):54-65. [doi: [10.1111/j.1365-277X.2012.01290.x](https://doi.org/10.1111/j.1365-277X.2012.01290.x)] [Medline: [23980931](https://pubmed.ncbi.nlm.nih.gov/23980931/)]
25. Thompson FE, Subar AF, Brown CC, Smith AF, Sharbaugh CO, Jobe JB, et al. Cognitive research enhances accuracy of food frequency questionnaire reports: results of an experimental validation study. *J Am Diet Assoc* 2002 Feb;102(2):212-225. [Medline: [11846115](https://pubmed.ncbi.nlm.nih.gov/11846115/)]
26. Dickin KL, Larios F, Parra PA. Cognitive Interviewing to Enhance Comprehension and Accuracy of Responses to a Spanish-Language Nutrition Program Evaluation Tool. *J Nutr Educ Behav* 2015;47(5):465-71.e1. [doi: [10.1016/j.jneb.2015.06.008](https://doi.org/10.1016/j.jneb.2015.06.008)] [Medline: [26209887](https://pubmed.ncbi.nlm.nih.gov/26209887/)]
27. Stopher P. Basic issues in surveys. In: *Collecting, managing and assessing data using sample surveys*. Cambridge: Cambridge University Press; 2012:64.
28. Ritchie J, Lewis J. Carrying out qualitative analysis. In: *Qualitative Research Practice - A Guide for Social Students and Researchers*. London: Sage Publications; 2003:220.
29. Magarey A, Watson J, Golley RK, Burrows T, Sutherland R, McNaughton SA, et al. Assessing dietary intake in children and adolescents: Considerations and recommendations for obesity research. *Int J Pediatr Obes* 2011 Feb;6(1):2-11. [doi: [10.3109/17477161003728469](https://doi.org/10.3109/17477161003728469)] [Medline: [20874449](https://pubmed.ncbi.nlm.nih.gov/20874449/)]
30. FNDE. RESOLUÇÃO/FNDE/CD/No 32 DE 10 DE AGOSTO DE 2006. 2006. URL: [ftp://ftp.fnde.gov.br/web/resolucoes\\_2006/res032\\_10082006.pdf](http://ftp.fnde.gov.br/web/resolucoes_2006/res032_10082006.pdf) [accessed 2016-11-03] [[WebCite Cache ID 6ljrYNUz9](#)]
31. Domel SB, Baranowski T, Leonard SB, Davis H, Riley P, Baranowski J. Accuracy of fourth- and fifth-grade students' food records compared with school-lunch observations. *Am J Clin Nutr* 1994 Jan;59(1 Suppl):218S-220S. [Medline: [8279428](https://pubmed.ncbi.nlm.nih.gov/8279428/)]
32. Hoelscher DM, Day RS, Kelder SH, Ward JL. Reproducibility and validity of the secondary level School-Based Nutrition Monitoring student questionnaire. *J Am Diet Assoc* 2003 Feb;103(2):186-194. [doi: [10.1053/jada.2003.50031](https://doi.org/10.1053/jada.2003.50031)] [Medline: [12589324](https://pubmed.ncbi.nlm.nih.gov/12589324/)]
33. Livingstone MB, Robson PJ, Wallace JM. Issues in dietary intake assessment of children and adolescents. *Br J Nutr* 2004 Oct;92(Suppl 2):S213-S222. [Medline: [15522159](https://pubmed.ncbi.nlm.nih.gov/15522159/)]
34. Burrows TL, Martin RJ, Collins CE. A systematic review of the validity of dietary assessment methods in children when compared with the method of doubly labeled water. *J Am Diet Assoc* 2010 Oct;110(10):1501-1510. [doi: [10.1016/j.jada.2010.07.008](https://doi.org/10.1016/j.jada.2010.07.008)] [Medline: [20869489](https://pubmed.ncbi.nlm.nih.gov/20869489/)]

*Edited by P Sullivan; submitted 11.08.15; peer-reviewed by D Bossen, M Raber; comments to author 23.09.15; revised version received 02.08.16; accepted 03.09.16; published 28.11.16.*

*Please cite as:*

*Fernandes Davies V, Kupek E, Faria Di Pietro P, Altenburg de Assis MA, GK Vieira F, Perucchi C, Mafra R, Thompson D, Baranowski T*

*Qualitative Analysis of Cognitive Interviews With School Children: A Web-Based Food Intake Questionnaire*

*JMIR Public Health Surveill* 2016;2(2):e167

URL: <http://publichealth.jmir.org/2016/2/e167/>

doi:[10.2196/publichealth.5024](https://doi.org/10.2196/publichealth.5024)

PMID:[27895005](https://pubmed.ncbi.nlm.nih.gov/27895005/)

©Vanessa Fernandes Davies, Emil Kupek, Patricia Faria Di Pietro, Maria Alice Altenburg de Assis, Francilene GK Vieira, Clarice Perucchi, Rafaella Mafra, Debbe Thompson, Thomas Baranowski. Originally published in JMIR Public Health and Surveillance (<http://publichealth.jmir.org>), 28.11.2016. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Public Health and Surveillance, is properly cited. The complete bibliographic information, a link to the original publication on <http://publichealth.jmir.org>, as well as this copyright and license information must be included.

Corrigenda and Addenda

# Metadata Correction of: How Best to Obtain Valid, Verifiable Data Online From Male Couples? Lessons Learned From an eHealth HIV Prevention Intervention for HIV-Negative Male Couples

Jason Mitchell<sup>1\*</sup>, MPH, PhD; Ji-Young Lee<sup>2\*</sup>, MSPH; Rob Stephenson<sup>3</sup>, PhD

<sup>1</sup>Office of Public Health Studies, University of Hawai'i at Manoa, Honolulu, HI, United States

<sup>2</sup>University of Miami Miller School of Medicine, Department of Public Health Sciences, Miami Beach, FL, United States

<sup>3</sup>Center for Sexuality and Health Disparities, Department of Health Behavior and Biological Sciences, University of Michigan School of Nursing, Ann Arbor, MI, United States

\*these authors contributed equally

**Corresponding Author:**

Jason Mitchell, MPH, PhD  
Office of Public Health Studies  
University of Hawai'i at Manoa  
1960 East-West Road  
Honolulu, HI, 96822  
United States  
Phone: 1 808 956 3342  
Fax: 1 808 956 5818  
Email: [jasonmit@hawaii.edu](mailto:jasonmit@hawaii.edu)

**Related Article:**

Correction of: <http://publichealth.jmir.org/2016/2/e152/>

(*JMIR Public Health Surveill* 2016;2(2):e155) doi:[10.2196/publichealth.6689](https://doi.org/10.2196/publichealth.6689)

The academic affiliation of author Jason Mitchell was incorrect in the paper entitled “How Best to Obtain Valid, Verifiable Data Online From Male Couples? Lessons Learned From an eHealth HIV Prevention Intervention for HIV-Negative Male Couples” (*JMIR Public Health Surveill* 2016;2(2):e152).

The affiliation of Jason Mitchell should be “Office of Public Health Studies, University of Hawai'i at Manoa, Honolulu, HI, United States”. The authors had brought this to the attention of JMIR editorial staff during proofreading; however, the error was only corrected in one of two instances where it needed to be edited.

In addition, author Jason Mitchell has also changed the phone number from 1 808-956-8577 to 1 808-956-3342.

Both these alterations have been made in the online version of the paper on the JMIR website on October 3, 2016 together with publishing this correction notice. Because these were made after submission to PubMed and other full-text repositories, the correction notice has been submitted to PubMed, and the original paper has been resubmitted to PubMed Central. The corrected metadata have also been resubmitted to CrossRef.

*Edited by G Eysenbach; submitted 23.09.16; this is a non-peer-reviewed article; accepted 23.09.16; published 26.09.16.*

*Please cite as:*

*Mitchell J, Lee JY, Stephenson R*

*Metadata Correction of: How Best to Obtain Valid, Verifiable Data Online From Male Couples? Lessons Learned From an eHealth HIV Prevention Intervention for HIV-Negative Male Couples*

*JMIR Public Health Surveill* 2016;2(2):e155

URL: <http://publichealth.jmir.org/2016/2/e155/>

doi: [10.2196/publichealth.6689](https://doi.org/10.2196/publichealth.6689)

PMID: [27694099](https://pubmed.ncbi.nlm.nih.gov/27694099/)

©Jason Mitchell, Ji-Young Lee, Rob Stephenson. Originally published in JMIR Public Health and Surveillance (<http://publichealth.jmir.org>), 26.09.2016. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Public Health and Surveillance, is properly cited. The complete bibliographic information, a link to the original publication on <http://publichealth.jmir.org>, as well as this copyright and license information must be included.



Corrigenda and Addenda

## Metadata Correction of: Lessons Learned From Methodological Validation Research in E-Epidemiology

Emmanuelle Kesse-Guyot<sup>1</sup>, PhD; Karen Assmann<sup>1</sup>, PhD; Valentina Andreeva<sup>1</sup>, PhD; Katia Castetbon<sup>2,3</sup>, PhD; Caroline Méjean<sup>1</sup>, PhD; Mathilde Touvier<sup>1</sup>, PhD; Benoît Salanave<sup>3</sup>, MD; Valérie Deschamps<sup>3</sup>, PhD; Sandrine Péneau<sup>1</sup>, PhD; Léopold Fezeu<sup>1</sup>, MD, PhD; Chantal Julia<sup>1,4</sup>, MD, PhD; Benjamin Allès<sup>1</sup>, PhD; Pilar Galan<sup>1</sup>, MD, PhD; Serge Hercberg<sup>1,4</sup>, MD, PhD

<sup>1</sup>Équipe de Recherche en Épidémiologie Nutritionnelle (EREN), Centre de Recherche en Épidémiologie et Statistiques, COMUE Sorbonne Paris Cité, Inserm (U1153), Inra (U1125), Cnam, Université Paris 13, Bobigny, France

<sup>2</sup>Ecole de Santé Publique, Centre de Recherche en Épidémiologie, Biostatistiques et Recherche Clinique, Université Libre de Bruxelles, Bruxelles, Belgium

<sup>3</sup>Unité de Surveillance et d'Épidémiologie Nutritionnelle (USEN), Institut de Veille Sanitaire, Centre de Recherche en Épidémiologie et Statistiques, COMUE Sorbonne Paris Cité, Université Paris 13, Bobigny, France

<sup>4</sup>Hôpital Avicenne, Département de Santé Publique, Bobigny, France

**Corresponding Author:**

Emmanuelle Kesse-Guyot, PhD

Équipe de Recherche en Épidémiologie Nutritionnelle (EREN)

Centre de Recherche en Épidémiologie et Statistiques

COMUE Sorbonne Paris Cité

Inserm (U1153), Inra (U1125), Cnam

Université Paris 13

74 rue Marcel Cachin

Bobigny, 93017

France

Phone: 33 0148388979

Fax: 33 0148388931

Email: [e.kesse@eren.smbh.univ-paris13.fr](mailto:e.kesse@eren.smbh.univ-paris13.fr)

**Related Article:**

Correction of: <http://publichealth.jmir.org/2016/2/e160/>

(*JMIR Public Health Surveill* 2016;2(2):e160) doi:[10.2196/publichealth.6876](https://doi.org/10.2196/publichealth.6876)

The authors of “Lessons Learned From Methodological Validation Research in E-Epidemiology” (*JMIR Public Health Surveill* 2016;2(2):e160) made a mistake in the name of author Katia Castetbon. The author's name should be Katia Castetbon instead of Katia Castebon.

This correction has been made in the online version of the paper on the JMIR Public Health and Surveillance website on October

31, 2016 together with publishing this correction notice. Because these were made after submission to PubMed and other full-text repositories, the correction notice has been submitted to PubMed, and the original paper has been resubmitted to PubMed Central. The corrected metadata have also been resubmitted to CrossRef.

*Edited by G Eysenbach; submitted 22.04.16; peer-reviewed by D Thompson, J Torous; comments to author 25.07.16; revised version received 25.08.16; accepted 28.09.16; published 31.10.16.*

*Please cite as:*

*Kesse-Guyot E, Assmann K, Andreeva V, Castetbon K, Méjean C, Touvier M, Salanave B, Deschamps V, Péneau S, Fezeu L, Julia C, Allès B, Galan P, Hercberg S*

*Metadata Correction of: Lessons Learned From Methodological Validation Research in E-Epidemiology*

*JMIR Public Health Surveill 2016;2(2):e160*

*URL: <http://publichealth.jmir.org/2016/2/e163/>*

*doi: [10.2196/publichealth.6876](https://doi.org/10.2196/publichealth.6876)*

*PMID: [27797322](https://pubmed.ncbi.nlm.nih.gov/27797322/)*

©Emmanuelle Kesse-Guyot, Karen Assmann, Valentina Andreeva, Katia Castetbon, Caroline Méjean, Mathilde Touvier, Benoît Salanave, Valérie Deschamps, Sandrine Péneau, Léopold Fezeu, Chantal Julia, Benjamin Allès, Pilar Galan, Serge Hercberg. Originally published in JMIR Public Health and Surveillance (<http://publichealth.jmir.org>), 31.10.2016. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Public Health and Surveillance, is properly cited. The complete bibliographic information, a link to the original publication on <http://publichealth.jmir.org>, as well as this copyright and license information must be included.

Original Paper

# Effect of Performance Feedback on Community Health Workers' Motivation and Performance in Madhya Pradesh, India: A Randomized Controlled Trial

Sangya Kaphle<sup>1</sup>, MA; Michael Matheke-Fischer<sup>2</sup>, MA; Neal Lesh<sup>1</sup>, PhD, MPH

<sup>1</sup>Dimagi Software Innovations, Cambridge, MA, United States

<sup>2</sup>Real Medicines Foundation, 11700 National Blvd, Suite 234, Los Angeles, CA, United States

**Corresponding Author:**

Sangya Kaphle, MA

Dimagi Software Innovations

585 Massachusetts Ave #3

Cambridge, MA, 02139

United States

Phone: 1 617 649 2214

Fax: 1 617 274 8393

Email: [sangyakaphle@gmail.com](mailto:sangyakaphle@gmail.com)

## Abstract

**Background:** Small-scale community health worker (CHW) programs provide basic health services and strengthen health systems in resource-poor settings. This paper focuses on improving CHW performance by providing individual feedback to CHWs working with an mHealth program to address malnutrition in children younger than 5 years.

**Objective:** The paper aims to evaluate the immediate and retention effects of providing performance feedback and supportive supervision on CHW motivation and performance for CHWs working with an mHealth platform to reduce malnutrition in five districts of Madhya Pradesh, India. We expected a positive impact on CHW performance for the indicator they received feedback on. Performance on indicators the CHW did not receive feedback on was not expected to change.

**Methods:** In a randomized controlled trial, 60 CHWs were randomized into three treatment groups based on overall baseline performance ranks to achieve balanced treatment groups. Data for each treatment indicator were analyzed with the other two treatments acting as the control. In total, 10 CHWs were lost to follow-up. There were three performance indicators: case activity, form submissions, and duration of counseling. Each group received weekly calls to provide performance targets and discuss their performance on the specific indicator they were allocated to as well as any challenges or technical issues faced during the week for a 6-week period. Data were collected for a further 4 weeks to assess intertemporal sustained effects of the intervention.

**Results:** We found positive and significant impacts on duration of counseling, whereas case activity and number of form submissions did not show significant improvements as a result of the intervention. We found a moderate to large effect (Glass's delta=0.97,  $P=.004$ ) of providing performance feedback on counseling times in the initial 6 weeks. These effects were sustained in the postintervention period (Glass's delta=1.69,  $P<.001$ ). The counseling times decreased slightly from the intervention to postintervention period by 2.14 minutes ( $P=.01$ ). Case activity improved for all CHWs after the intervention. We also performed the analysis by replacing the CHWs lost to follow-up with those in their treatment groups with the closest ranks in baseline performance and found similar results.

**Conclusions:** Calls providing performance feedback are effective in improving CHW motivation and performance. Providing feedback had a positive effect on performance in the case of duration of counseling. The results suggest that difficulty in achieving the performance target can affect results of performance feedback. Regardless of the performance information disclosed, calls can improve performance due to elements of supportive supervision included in the calls encouraging CHW motivation.

(*JMIR Public Health Surveill* 2016;2(2):e169) doi:[10.2196/publichealth.3381](https://doi.org/10.2196/publichealth.3381)

**KEYWORDS**

community health workers; performance feedback; motivation; supportive supervision; mHealth apps

## Introduction

The success of small-scale community health worker (CHW) programs in providing basic health services and strengthening health systems in resource-poor settings is well documented in the literature [1-4]. CHWs operating in small-scale, well-managed projects can be effective agents of change, but often even small-scale programs lack adequate focus, reporting and documentation, training, monitoring, and supervisory and support mechanisms to encourage and motivate the CHWs to excel at their jobs [5,6].

Health worker motivation is identified as an important determinant of performance by many studies [7-10]. Other than financial incentives, important determinants of performance and motivation for CHWs include work conditions, training, audit and feedback, reminders, and supportive supervision [7,8]. Multifaceted interventions that target more than one of these factors are more likely to improve performance than single interventions [8]. Appreciation by managers, colleagues, and the community, as well as a stable job, income, and training are additional motivating factors for CHWs [9,11-13].

Supportive supervision targeting health workers' knowledge and skills, motivation, and adherence to correct practices provide incentives that positively impact performance [8,11]. In addition to performance monitoring and providing feedback, providing health workers with specific targets can also improve CHW performance. Opening two-way communication channels reduces the sense of isolation that many CHWs face while working in remote areas and can increase internal motivation, positively impacting health worker performance and quality of care [8].

In this study, we focus on improving CHW performance by providing individual feedback on specific indicators to CHWs working in an international nongovernmental organization-run mHealth program to address malnutrition in children younger than 5 years. We open two-way communication channels by placing weekly calls to CHWs to provide feedback on their performance and discuss any challenges they are facing in their work. The aim of the intervention is to gain improvement in the performance indicators, through improvement in CHW motivation, knowledge, skills, and adherence to reporting standards. With improved ability to monitor performance and provide targeted support to CHWs, we hope to improve their motivation and performance.

## Analytical Framework

### *Intrinsic and Extrinsic Motivation*

Many theories of motivation distinguish between intrinsic and extrinsic work motivation. Intrinsic motivation involves people deriving satisfaction from doing the activity because they find the activity itself rewarding [14]. On the other hand, extrinsic motivation requires "an instrumentality" between the activity

and a separable consequence, such as tangible or verbal rewards [14,15].

In most work contexts, both extrinsic and intrinsic motivations come into play and affect worker performance. Although early research supported the additivity of intrinsic and extrinsic motivation, more recent work has also highlighted the tradeoffs between the two, with external rewards designed to enhance external motivation detracting from internal motivation [16,17]. Verbal rewards and feedback are found to add to intrinsic motivation compared with more tangible rewards [18-20].

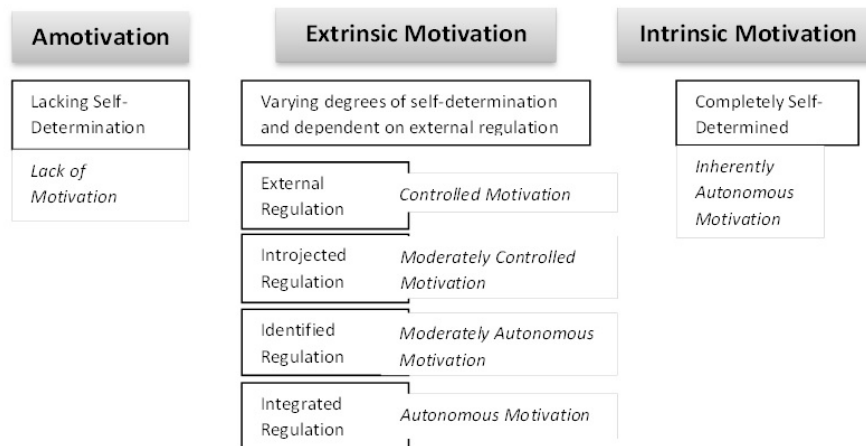
### *Self-Determination Theory*

The complex relationships between intrinsic and extrinsic motivation is best explained by self-determination theory (SDT), which distinguishes between autonomous motivation, analogous to intrinsic motivation, and controlled motivation, analogous to extrinsic motivation. Any behavior falls somewhere on the continuum of controlled to autonomous motivation [14]. A behavior fueled by external motivators can undergo a process of internalization, whereby it is "taken in" and becomes autonomous to the individual [21].

Figure 1 shows the internalization continuum described by SDT. Adapted from Gagne and Deci [22], Figure 1 describes the self-determination continuum of motivation from amotivation where the individual is completely lacking in motivation and there are no intentional regulations present, degrees of extrinsic motivation where contingencies and rewards are present and internalized to varying degrees resulting in degrees of self-determination, to intrinsic motivation where the individual's motivation is completely self-determined. It shows four different levels of internalization described by the theory: (1) external regulation, in which the regulation has not been internalized and depends on contingencies of rewards or punishment; (2) introjection in which a regulation has been taken in, but has not been accepted as the person's own and it is still controlling the person's behavior; (3) identification, in which the regulation is more in-line with the person's values, personal goals, and identities, and the behavior has an internal perceived locus of causality; and (4) integrated regulation in which people have the sense that the behavior is more central to their identity, it emanates from their sense of self, and is seen as self-determined and autonomous [14].

SDT postulates that increased autonomy, competence, and relatedness are the three factors or "needs" that underlie intrinsic motivation and people need to feel autonomous, competent, and need to relate to the work environment to maintain their intrinsic motivation. When all three needs are satisfied, integration is more likely to be achieved. Of these factors, autonomy support, including choice, feedback, and positive interpersonal climate with open communication and empathy, has been identified as more important for internalization [23,24]. Full internalization of external motivation has shown to yield increases in performance and work outcomes [14,25].

**Figure 1.** Self-determination continuum of motivation.



### Goal-Setting Theory

Locke et al’s [26] goal-setting theory posits that goals are effective in enhancing performance when (1) they are specific, optimally difficult, and have high valence; and (2) workers understand what behavior is needed to achieve those goals and feel competent to carry out those behaviors. Goals affect performance through four mechanisms: (1) they direct effort toward goal-relevant activities and away from irrelevant activities, (2) they have an energizing function and higher goals draw greater efforts, (3) they affect persistence, and (4) they affect action indirectly by leading to discovery or use of task-relevant knowledge and skills [27].

Ensuring that the goals are viewed as important and increase self-efficacy are the two important factors that lead to goal commitment and acceptance. Goal commitment is important and the goal-performance relationship is strongest when individuals are committed to their goals. Feedback is another important factor and it is important to know one’s progress against a target to adjust the level of effort or performance strategies to match the requirements of the target. Many studies have found that a combination of goals and feedback is more effective than goals alone [28-32].

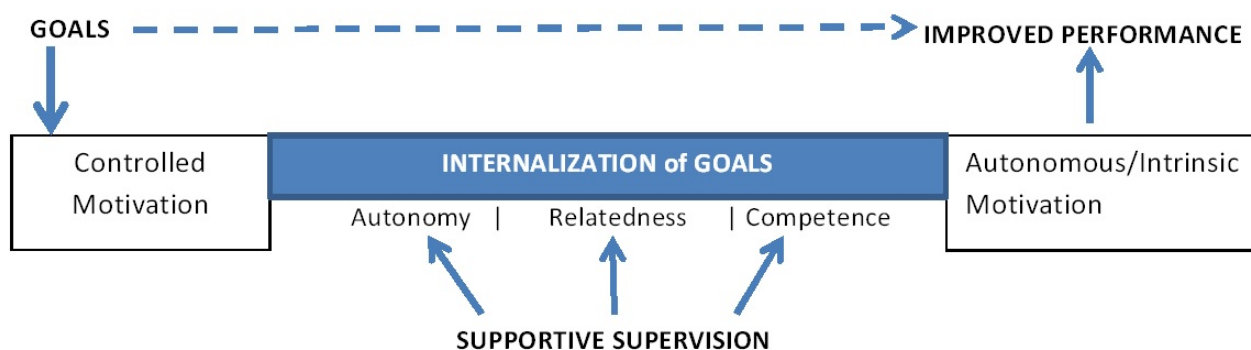
Many studies set in different countries and contexts have established that the relationship between goal setting and increased performance is reliable, although some negative effects of goal setting, such as fraudulent reporting, are possible outcomes [33].

### Community Health Worker Motivation, Goal Setting, and Performance

Our analytical framework rests on SDT and goal-setting theory, whereby organizational goals and behavioral standards are internalized by the CHW, enhancing autonomous motivation and performance. We expect that by providing targets, performance feedback, and supportive supervision through weekly phone calls, we will influence the three factors (autonomy, competence, and relatedness) that lead to greater internalization of the external regulation and goals and this will thus increase CHW performance and it will be sustained over time.

Figure 2 presents the conceptual framework used to develop our intervention. Since goal setting alone can lead to increased performance [26], CHWs are provided with performance targets. Supportive supervision that makes health workers feel cared for and provides recognition, appreciation, and encouragement supports autonomy in the CHW and enhances intrinsic motivation [34]. The CHWs working in remote contexts where interactions with peers and the organization is limited will have a greater sense of relatedness with regular two-way communication with someone they perceive to be higher up in the organizational structure. Setting clear goals and targets and providing performance feedback and need-based training to improve gaps in performance, improve knowledge, and set clear guidelines for practices and reporting standards can also be effective in promoting internalization and improving CHW self-efficacy, competence, and performance. A combination of goal setting and a performance feedback loop providing supportive supervision to CHWs works as a multifaceted incentive aimed at improving CHW motivation.



**Figure 2.** Conceptual framework: community health worker motivation, goal setting, and performance.

### ***Supportive Supervision, Motivation and Performance of Community Health Workers***

Regular and reliable supervision is cited as a missing link in increasing CHW performance in India [35]. Supportive supervision, community recognition and respect, peer support and learning, community information systems, and having clear roles, responsibilities, and targets are among other nonmonetary factors that incentivize CHW performance [34-41].

Appreciation and encouragement are important elements of supportive supervision. Empirical studies find positive effects of verbal rewards on intrinsic motivation and performance [42,43]. Other recent studies also identify lack of appreciation and nonrecognition of performance as demotivators for health workers [11-13]. Communicating goals clearly and making health workers feel supported and cared for are also deemed important for motivation [44].

### **The Project in Madhya Pradesh, India**

Real Medicine Foundation (RMF) has been running a mobile health project using CommCare, with 60 community nutrition experts, targeting infant malnutrition in 600 villages located in five districts of Madhya Pradesh, India: Jhabua, Alirajpur, Barwani, Khandwa, and Khargone. Their aim is to increase communities' awareness about malnutrition, mobilize communities to increase demands for treatments and services to alleviate malnutrition, and strengthen the capacity of frontline workers such as Anganwadi workers in combatting malnutrition. The community nutrition experts have been working since 2010 and the program has been in effect with CommCare since July 2011.

Since the program began, RMF has successfully treated 2157 children at nutrition rehabilitation centers, improved the nutritional status of 24,822 children suffering from moderate

and severe acute malnutrition, and trained more than 329,780 individuals on malnutrition prevention, awareness, and treatment. RMF works closely with the Government of Madhya Pradesh.

The community nutrition experts in the RMF program are monitored and supervised primarily via their district supervisor. Although the community nutrition experts do receive some feedback from their district supervisors, they do not receive any individual-level feedback on performance metrics. The community nutrition experts attend a meeting at the district headquarters every 15 days. The district supervisor reviews the reports the community nutrition experts submit on child health status, discusses issues that the community nutrition experts have faced in the field, and develops a roadmap for the next 15 days. There is general acknowledgment of the community nutrition experts who have performed well for that period. The district supervisor also gives specific acknowledgment to community nutrition experts when they excel in the group environment to motivate other community nutrition experts to improve their performance. The district supervisor also accompanies the community nutrition experts on home visits to monitor performance. She spends the entire day with the community nutrition expert covering the households she visits and monitors the community nutrition expert's work closely. She provides individual feedback/support at this time, which includes supporting community nutrition experts to motivate households that are resistant to admitting their child to the nutrition rehabilitation centers and supervising reporting (ie, reviewing how the data are entered on CommCare or their paper-based forms).

### ***Motivators for the Community Nutrition Experts***

Focus group discussions and unstructured interviews with the community nutrition experts were conducted to gain insights on their goals, motivation, challenges, and support structures

affecting their work. We conducted two focus group discussions with three and 10 community nutrition experts, respectively, and interviewed five community nutrition experts and all five district supervisors to gain better understanding of what motivates these health workers. Field observations shadowing five community nutrition experts also informed understanding of the community nutrition experts work structure and use of their mobile app.

The motivation goals of the community nutrition experts seem to be rooted in experience and the contributions they make to their households as a result of their job. Most community nutrition experts did not have prior work experience and most mentioned that they could not imagine going back to the situation when they did not work. They derive intrinsic motivation from their professional conscience and the visible improvements they see in the nutritional status of children in their catchment area. Working in the communities for 2 to 3 years, they form good relationships with the villagers. As a result, a negative outcome, such as the death of a patient, can be demotivating for them. The financial contributions they make to their households are a major motivational factor because their salary has made a significant impact on their household living standards. Their experiences have made them more confident and eager to learn more, and in many cases enlisted a desire to progress further and deal with more responsibility and gain higher salaries, although most do not have any set goals toward this objective. Very few community nutrition experts have set life or career goals for themselves. Their motivation seems to be based on a combination of intrinsic and extrinsic factors.

## Methods

In this study, we examined the impact of performance feedback on CHW performance and motivation in a randomized controlled trial with a sample of 60 CHWs who were part of the community nutrition program with RMF in Madhya Pradesh, India.

## Study Design

### *The Intervention*

In their article addressing ways to improve CHW performance in India, Bajpaai and Dholakia [35] suggested additional monitoring done with minimal paperwork to improve performance of accredited social health activists. Feedback mechanisms, such as text messages or telephone complaint services, were also suggested as viable channels to provide and receive efficient feedback.

To communicate detailed performance feedback, we set up a call center, placing weekly calls to the community nutrition experts and relaying feedback on performance metrics. Phone calls provided a way to discuss and receive feedback from the community nutrition experts regarding any work-related issues, personal needs, mistreating, or technical difficulties they might have been having with the mobile app, which could be escalated to district supervisors so that they could be resolved quickly.

The European Foundation of Quality Management identifies the following human resource management principles as effective motivators for health works: supervision schemes, recognition schemes, performance management, training and professional development, leadership, participation mechanisms, and intraorganizational communication processes [44]. These motivators are effective in promoting internalization of controlled motivation by supporting perceptions of autonomy, relatedness, and competence experienced by health workers. Our intervention to provide performance feedback to community nutrition experts factored in six of these seven criteria. The community nutrition experts received supportive supervision, recognition for their work, performance feedback, training in problem areas regarding use of their app, and they had an opportunity to contribute to the program by discussing problems and challenges. Any issues identified during the calls were escalated to the program coordinator, thus contributing to intraorganization communication. Table 1 describes elements of the intervention and how they were related to supporting autonomy, competence, and relatedness to facilitate internalization and achieve sustained behavior change in the absence of the intervention.

**Table 1.** Elements of the intervention and the factors supporting internalization.

Elements of the intervention	Internalization factors	Rationale
Greetings, introductions, discuss well-being, ask if good time to call	Relatedness, autonomy	Discussing well-being of the community nutrition expert and family displays concern and empathy, and supports autonomy and relatedness because the community nutrition expert can voice any personal difficulties affecting her work.
Targets and performance feedback	Competence, autonomy	Providing performance feedback against goals targets competence and autonomy by discussing gaps in performance and proving the community nutrition expert has the means to assess her own performance as well.
Congratulations/encouragement	Autonomy	Encouragement and positive feedback increase autonomy by instilling a sense of self into the community nutrition expert's work.
Reminders and retraining on app or work flow	Competence	Retraining and reminders on the community nutrition expert's desired workflow or on how to use the app increases their competence.
Discuss work challenges and ways to support their work	Autonomy, relatedness	Participation and being able to discuss any challenges or issues will increase autonomy by empowering the community nutrition expert to provide input. Discussing concerns of others improves relatedness.
Technical support regarding phone or app	Relatedness, competence	Any technical problems arising on the phone can be fixed and community nutrition expert will be better able to perform the job. Discussing these issues will also improve relatedness
Reminder of target and next call, goodbyes	Autonomy, relatedness	We provide targets with the aim to improve autonomy by giving the community nutrition expert the information to do her work. It could seem like external pressure, although it is communicated positively. Warm, encouraging, interpersonal communication supports relatedness and autonomy.

### Performance Indicators

Three performance indicators and respective targets were identified by the RMF program as relevant indicators to measure the performance of the community nutrition experts: (1) case activity, (2) number of infant and child health form submissions, and (3) duration of counseling.

### Case Activity

Case activity was defined as the number of clients out of total clients that were visited in the last 2 weeks. The target was to visit all the registered clients every two weeks. By the end of each 2-week period, each community nutrition expert's case activity was expected to be 100%, regardless of the number of clients registered. The case activity metric monitored the community nutrition expert's coverage of the villages and clients and provided a measure for access to care available to the communities they served. Goal-setting theory suggests that goal difficulty has an inverse relationship with effort, with moderately difficult goals drawing the most effort out the individual and easy and very difficult goals drawing less effort and motivation to meet the target [27]. Because the community nutrition experts had different numbers of clients, the degree of difficulty of achieving this target was not consistent across the 60 community nutrition experts.

### Form Submissions

The form submissions indicator was the total number of infant health and nutrition forms submitted in the last week. This form collects information on the child's nutritional and health status and tracks their nutritional progress at each follow-up visit by the community nutrition expert. The target here was for each community nutrition expert to fill out this form each time she visited a household to register or follow up with a child. Hence, the number of form submissions was measured against the total number of clients visited in the last week (case activity), which

provided the target for this particular indicator. The community nutrition expert could exceed the target on this metric because case activity did not register repeat visits within a period as a new visit or account for more than one form submitted during a visit. Hence, the number of form submissions for the infant health and nutrition form could be higher than the case activity if the community nutrition expert visited the same case more than once during the week. The target for form submissions was measured based on case activity; the community nutrition expert was expected to fill out at least as many infant health and nutrition forms as the number of clients visited in the last week.

### Duration of Family Counseling

Family counseling duration was the adjusted mean time spent counseling families on infant nutrition in the last week. The target was for each community nutrition expert to spend at least 15 minutes counseling each family on the topics related to mother and infant nutrition. Tracking time taken by each community nutrition expert to counsel families on the importance of hygiene, nutrition, and admitting malnourished children to the nutrition rehabilitation centers provided a measure for the quality of care available to the communities. Because we wanted to ensure that each family received this minimum amount of counseling, all counseling times longer than 20 minutes were rounded down to 20 minutes during the data collection stage before any analysis. This was done in consultation with RMF, which had a target of 15 minutes for each counseling session. Tracking counseling times was made simple with CommCare, which records the time taken to complete each form and the community nutrition experts use a family counseling form to counsel the families.

### **Randomization and Intervention Stages**

The 60 community nutrition experts were randomized into three treatment groups, one for each of the three performance indicators (case activity, form submissions, duration of counseling). The study period was 10 weeks in total and was divided into two stages: the intervention period and the postintervention period. In the intervention period, which lasted for 6 weeks, each community nutrition expert received a weekly phone call discussing their performance on a certain indicator against its target. During this period, the other two groups served as the control group. We had imbalanced randomization for the analysis (2:1) and each treatment had 20 community nutrition experts; therefore, each control had 40 community nutrition experts.

In the postintervention period, which lasted for 4 weeks, the community nutrition expert did not receive any feedback to examine whether the effects of the feedback in the intervention period were sustained or wore off over time. Because we were interested in studying dynamic, intertemporal effects of our intervention, this was a crucial time period for our study in which we could examine whether any effects on performance and reporting standards were lasting. For instance, if we found that results were positive and sustained over the two study periods, the effects of the intervention could be deemed more effective than if these effects were attributed to a novelty factor and did not carry over time without constant monitoring and feedback. We did not have a control group that did not receive any calls because the RMF program requested that all community nutrition experts receive some feedback on their performance.

Randomly allotting the community nutrition experts into the three treatment groups could result in lopsided treatment groups based on baseline performance. To create balanced treatment groups, we first ranked the community nutrition experts baseline performance on each of the three indicators. We gave a percentage score for each indicator by comparing the community nutrition expert's performance on each indicator against the target for that indicator and converted these into  $z$  scores so that they were normalized. We then tallied up baseline performance scores for the three indicators using the  $z$  scores for each community nutrition expert and used this score to rank the community nutrition expert's baseline performance. For randomization, the community nutrition experts were sorted by baseline rank in groups of three. To generate balanced treatment groups in terms of baseline performance, we randomly allocated each community nutrition expert in the group to the three treatment groups.

### **Data**

Data for the study on the three indicators were obtained directly from CommCareHQ, CommCare's cloud-based server where data recorded by the community nutrition expert were stored in real time. Additional data on the community nutrition expert's demographic characteristics, personal traits, and motivation were collected in a survey conducted at the end of the postintervention period. We collected additional qualitative data from interviews with five community nutrition experts and the district coordinators as well as a focus group discussion with

10 community nutrition experts exploring motivations and challenges faced by community nutrition experts in their work and communities at the start of the study. We also collected some qualitative information at the end of the study to follow up on what was driving the results. This was done by randomly selecting three community nutrition experts from each treatment group for follow-up interviews to assess the main drivers of any changes in performance from their perspective.

We had panel data spanning 10 weeks, in which we collected performance data on each indicator for the three treatment groups. Our dataset had some important limitations, the more pressing of which was measurement error due to technical errors arising in the app. The community nutrition experts were working in areas with low network connectivity and were using an older version of the app, which had some technical errors that were not resolved before the start of our intervention. Working in low network connectivity means that the telecommunications network required to send the collected data to the central server was often inadequate or missing to send the information; therefore, there was measurement error in the data we could access from the server. Other technical errors, such as app error or missing multimedia and other bugs, meant that sometimes community nutrition experts were unable to access the app to record and send the data to the server. Many community nutrition experts also had repeat registrations for the same client with up to five repeat entries. This also affected our data quality for the first indicator—case activity—because this was measured against the total number of registered clients. During the intervention, community nutrition experts were advised not to fill out repeat registrations as they were doing before the intervention, meaning that the number of form submissions could decrease following the intervention. Similarly, we did not have data for total clients visited for the postintervention period, meaning that we did not have the target against which to measure form submissions data for that period. Therefore, form submissions is divided into two indicators: (1) form submissions number (ie, number of forms submitted in the last week), which was analyzed in the intervention and postintervention period; and (2) form submissions proportion, the number of forms as a proportion of total clients visited in the last week, which was analyzed only in the intervention period.

### **Qualitative Data**

We administered an end-line survey to all community nutrition experts who were not lost to follow-up to capture their perception of the calls. We also randomly selected three community nutrition experts from each of the three treatment groups and conducted unstructured interviews on the phone to understand what was driving the results. The interviews were conducted by the program coordinator for RMF's nutrition program.

We also conducted two focus group discussions during the intervention with three and 10 community nutrition experts, respectively, and interviewed five community nutrition experts and all five district supervisors to gain a better understanding of what motivated these health workers. Field observations shadowing five community nutrition experts also informed



understanding of the community nutrition experts work structure and use of the mobile app. No feedback calls were placed during this time to any of the community nutrition experts.

Our aim was to study the effects of providing performance feedback on CHW motivation and performance. Because motivation is difficult to measure directly, we focused on assessing the effects of the feedback calls on CHW performance on specific indicators. The effects of other motivators could not be analyzed directly because of the study design, which did not include a control group that did not receive any calls. Therefore, the main focus of the analysis and empirical strategy is on the effects of the performance feedback.

The indicators used in this study were identified as direct measures of quality and access to health care because they were important for the RMF program and for other programs using mHealth. Adherence to correct practices in reporting and utilization of the mHealth platform is important and cannot be separately measured from the performance indicators discussed subsequently because all data were collected using CommCareHQ, which does not distinguish between reporting and performance. Accurate reporting is a vital element of the community nutrition expert's performance and we viewed any change in performance indicators due to adherence to reporting standards as an improvement in CHWs overall performance as well.

### **Feedback Procedure**

The call center agent for our intervention was an anonymous call center operator, who was introduced to the community nutrition experts as a member of CommCare by RMF before the start of the intervention. The community nutrition experts were told to expect calls from CommCare and respond to the call by the RMF program. Due to unforeseen personal reasons, the call center operator changed after 3 weeks of the intervention and the researcher (SK), also introduced as a member of CommCare, placed the remainder of the calls. The community nutrition experts had no prior relationship to the call center operator or the researcher before the study. All the community nutrition experts received face-to-face interaction at the end of the study period for a survey and 18 of 60 community nutrition experts received additional visits, 13 in the form of a focus group discussion and five in the form of shadow visits to their clients during the intervention.

We began each call by greeting the community nutrition expert and asking her about her family and well-being; we then asked if she had some time to discuss her performance for that week. The community nutrition experts then received their performance feedback. The feedback was nuanced depending on whether she (1) met/exceeded her target, (2) improved a lot or little, or (3) showed no improvement in her work. All feedback included a reminder of the target, appreciation for their efforts, and a discussion of problems and solutions regarding their work. Those with little or no improvement were encouraged to meet their target and retrained on using the app and the desired workflow regarding the relevant indicator. Next, technical problems in the app were discussed and recorded to be escalated to the project coordinator. Finally, community

nutrition experts received another reminder about their targets, the day and time of the next phone call, and a warm goodbye.

The mean duration of the calls was approximately 5 minutes, although the initial calls were longer. Each community nutrition expert received a call on a particular day of the week at a particular time determined by the community nutrition expert. Because the community nutrition experts lived in areas with poor network connectivity and could be busy with other work, we attempted to call every community nutrition expert five times in case she could not respond or the call did not go through. Three of these attempts were made on the day the community nutrition expert was scheduled to receive the call, one was made the following day, whereas those who were not reached during the week received an additional attempt at the end of the week when they were not at work.

### **Empirical Strategy**

We estimated main treatment effects and heterogeneous effects of call intensity (number of calls received) for each of the three performance indicators using a random effects model. We first estimated double difference estimates for main effects and heterogeneous effects in the intervention period, and then expanded the analysis to the postintervention period to estimate sustained effects of the intervention. The equations used to look for the main effects of the treatment for each indicator (double difference estimator, cross-partial effects, and comparing all three indicators head-to-head) as well as the equation used to estimate heterogeneous effects are presented in [Multimedia Appendix 1](#).

### **Statistical Tests**

We used a range of statistical tests fitting the nature of the variables. Community nutrition expert characteristics were compared using one-way ANOVA because we were comparing the means of three different treatment groups, whereas the Fisher exact test was used for the categorical variables because of the small number of observations in some of the categories. Binomial tests were used for the binary variable "other work." Because the performance indicators did not follow a normal distribution, following Gneezy and Rustichini [45], we used the Mann-Whitney *U* test to compare the medians of the three treatment groups to one another. We also used unpaired *t* tests with unequal variance to compare means of the treatment group with the relevant control group, and the Fisher exact test for categorical variables.

### **Ethical Considerations**

We took all measures possible to ensure that our study followed research governance and ethical protocols necessary for such research. The study benefited the participating community nutrition experts and the larger community where they worked by exploring ways to improve the quality and experience of the health services they provided to their communities. The study posed minimal risk to the participants and was part of the normal operational experimentation done with CommCare users. The community nutrition experts often received feedback through their district supervisors, which is in-line with the feedback provided in the intervention. There was no consequence to the community nutrition expert's job as a result of the experiment

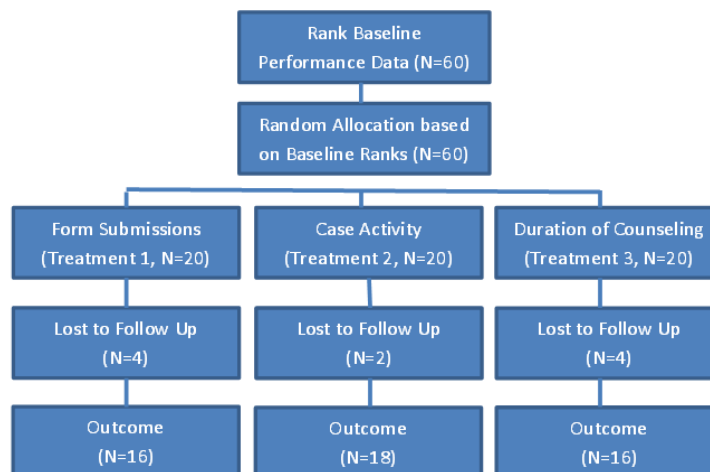


because we did not systematically share performance data with RMF program managers making hiring and firing decisions.

## Results

In total, 60 community nutrition experts were randomized into three treatment groups based on baseline performance. Figure 3 shows the flow of participants through the study.

**Figure 3.** Randomization and treatment allocation of participants.



**Table 2.** Descriptive statistics, community nutrition expert characteristics, and performance indicators.

Characteristics and indicators	Total (N=60)	Treatment group		
		Form submission (n=20)	Case activity (n=20)	Duration of counseling (n=20)
<b>Community nutrition expert characteristics, mean (SD)</b>				
Baseline rank	30.4 (17.22)	29.95 (17.61)	30.6 (17.29)	30.65 (18.03)
Age (years)	32.15 (7.49)	31.21 (5.84)	32.15 (6.81)	33.05 (9.54)
Number of children in household	1.73 (1.14)	2.05 (1.02)	1.85 (1.18)	1.3 (1.13)
Number of adults in household	3.63 (1.98)	3.42 (1.92)	3.35 (2.03)	4.1 (1.99)
Education (years)	11.55 (2.59)	10.57 (2.24)	11.8 (2.59)	12.25 (2.73)
<b>Performance indicators,<sup>a</sup> mean (SD)</b>				
Form submissions (n)	43.68 (48.39)	41.75 (47.69)	42.4 (50.63)	46.9 (49.15)
Form submissions (proportion)	0.42 (0.43)	0.41 (0.45)	0.37 (0.37)	0.49 (0.48)
Case activity	17.65 (24.55)	18.33 (27.00)	16.4 (22.99)	18.20 (24.71)
Duration of counseling (mins)	2.01 (3.24)	2.44 (4.31)	1.92 (2.69)	1.67 (2.57)

<sup>a</sup> Form submissions number was measured as the number of form submissions in a week. Form submissions proportion was measured as form submissions as a proportion of total clients. Case activity was measured as percentage of total clients visited in a 2-week period. Duration of counseling was measured as duration of family counseling in minutes.

The mean age of the community nutrition experts was 32 (SD 7.49) years with mean 11.5 (SD 2.59) years of education. Approximately 24% (14/60) of the community nutrition experts were engaged in other work, including agricultural work, in addition to their job with RMF. Baseline performance indicator means for the community nutrition experts were (1) form submissions (mean 43.68, SD 48.39), (2) case activity (mean 17.65, SD 24.55), and (3) duration of counseling (mean 2.01, SD 3.24). We did not find any significant differences between the three treatment groups in terms of rank, age, education,

## Descriptive Statistics and Baseline Comparisons

Table 2 presents summary statistics of some community nutrition expert characteristics and baseline performance indicators of the entire sample and by treatment groups.

number of children, number of adults, marital status, and year started as community nutrition expert. Similarly, baseline performance indicators also did not exhibit a difference in medians for the three treatment groups.

The treatment groups and their relevant control group were also balanced in terms of community nutrition expert characteristics and performance indicators. However, we did see some significant differences between treatment and control means

for duration of counseling in terms of number of children ( $P=.04$ ) and education (Fisher exact= $0.055$ ).

### Differential Attrition

We tested for differential attrition across the entire sample and within the three treatment groups. The attrition rate for the entire sample was high at 17% (10/60) and 10 of 60 community nutrition experts were lost to follow-up during the course of the study. Most of these were due changes in the community nutrition expert appointments made by the RMF program during the course of the study. We had differential attrition based on rank in our data because (1) the program replaced some lower-performing community nutrition experts during the course of the study (this was done outside of our intervention because the performance feedback was not shared with RMF program managers) and (2) the data in CommCareHQ was not updated at the start of the study, so some community nutrition experts who dropped out or were replaced before the start of the study were included in the baseline data, whereas their replacements were included in the end-line data (ie, we did not have baseline data for the replacements). We found differential attrition for baseline rank ( $P<.001$ ) with lower-ranked community nutrition experts dropping out more than those with a higher ranking in terms of baseline performance indicators. We also found differential attrition significant for age ( $P=.009$ ), with younger community nutrition experts dropping out, and year started as a community nutrition expert ( $P=.003$ ).

We found differential attrition in terms of rank for all three treatment groups and in terms of age ( $P=.02$ ) and year started as community nutrition expert ( $P=.04$ ) only for those in in the duration of counseling treatment group. In terms of rank, we found significant differential attrition for duration of counseling ( $P=.01$ ). The form submission group ( $P=.09$ ) and the case activity group ( $P=.09$ ) did not have significant differential attrition for rank. There was no differential attrition effect for education, number of children, and number of adults, marital status, or other work for all three intervention groups.

Lower-performing community nutrition experts in the baseline period who dropped out of the study in all three treatment groups could have affected our results.

To ensure that results were not driven as a result of the differential attrition, we re-estimated equations 1 and 4 in [Multimedia Appendix 1](#) to account for attrition biases by matching the community nutrition experts lost to follow-up with those with the closest rank in the same treatment group. [Multimedia Appendix 2](#) presents the results of the matching and shows the rank, age, education, marital status, number of adults, number of children, start year, and other work data for the community nutrition experts lost to follow-up and their replacements.

### Main Effects

[Tables 3](#) and [4](#) present the main treatment effects from equation 1 ([Multimedia Appendix 1](#)) in the intervention period and the postintervention period to identify intertemporal, sustained effects. There are three different regression specifications for each of the three indicators. The variables were defined so that case activity, form submissions, or duration of counseling indicated the treatment group, with a value of 1 if the community nutrition expert belonged to that treatment group and 0 otherwise. The variable “after” indicated the time variable with a value of 0 if the data were from the baseline period and 1 if from the intervention or postintervention period. The variable “treatment×after” indicated our double difference estimator and was an interaction term between the treatment and time variables. Form submissions were analyzed in terms of absolute number of form submissions and proportion of form submissions to total clients visited in a 7-day period. The results to correct for differential attrition were also included in [Tables 3](#) and [4](#), and show the double difference estimates after matching and replacing community nutrition experts lost to follow-up with those with the closest baseline rank in the same treatment group. The estimated coefficients give percentage point changes for case activity, form submissions proportions and absolute changes (ie, change in absolute number of form submissions), and absolute number of minutes for duration of counseling.

**Table 3.** Difference in differences (DID) estimates: impact of calls on performance indicators during the intervention stage.

Dependent variable	Case activity			Form submissions (number)			Form submissions (proportion)			Duration of counseling		
	DID	<i>t</i> ( <i>df</i> )	<i>P</i>	DID	<i>t</i> ( <i>df</i> )	<i>P</i>	DID	<i>t</i> ( <i>df</i> )	<i>P</i>	DID	<i>t</i> ( <i>df</i> )	<i>P</i>
<b>Original sample (n=300)</b>												
Treatment	-3.98	-0.59 (299)	.55	-7.09	-0.97 (299)	.33	-0.042	-0.30 (299)	.76	-0.54	-0.37 (299)	.71
After	9.27	2.65 (299)	.008	-36.94	-8.93 (299)	.001	0.22	3.31 (299)	.001	-0.09	-0.11 (299)	.91
Treatment×after	4.66	0.80 (299)	.42	2.89	0.40 (299)	.68	-0.002	-0.02 (299)	.98	3.86	2.91 (299)	.004
Constant	22.20	5.46 (299)	<.001	52.15	12.57 (299)	.001	0.49	6.27 (299)	.001	2.43	2.97 (299)	.003
<b>Correcting for attrition (n=360)</b>												
Treatment	-6.57	-1.48 (359)	.13	-1.26	-0.31 (359)	.76	-0.12	-1.35 (359)	.17	-0.64	-0.49 (359)	.62
After	12.43	4.41 (359)	<.001	-29.50	-8.80 (359)	.001	0.26	4.44 (359)	.001	0.07	0.1 (359)	.92
Treatment×after	4.78	1.23 (359)	.21	-1.57	-0.40 (359)	.68	0.04	0.49 (359)	.62	3.48	2.88 (359)	.004
Constant	21.78	5.28 (359)	.001	44.58	10.97 (359)	.001	0.51	5.91 (359)	.001	2.31	3.06 (359)	.002

The results showed a significant increase in performance for duration of counseling in the intervention period, whereas there were no significant treatment effects for case activity and form submissions. The interaction of counseling×after, which captured the treatment effect for receiving feedback on the duration of counseling, was significant at the 90% confidence interval, and receiving performance feedback on the duration of counseling increased mean counseling times by 3.860 minutes ( $P=.004$ ), which had an effect size of Glass's  $\delta=0.974$  (control group SD 2.910) and Cohen's  $d=0.778$  (pooled SD 4.976), both of which indicate a moderate to large effect of receiving duration of counseling feedback. After correcting for attrition bias, the effect size was Cohen's  $d=0.710$  (pooled SD 4.908) and Glass's  $\delta=0.883$  (control SD 3.945).

In Table 4, analyzing the intervention and postintervention period data together, we found that the impact of receiving performance feedback on duration of counseling on mean counseling times was sustained. Counseling×after, which captured the treatment effect for receiving performance feedback

on the duration of counseling during the intervention and postintervention periods, was significant and counseling times increased by a mean 4.469 minutes ( $P=.004$ ) when we included the postintervention data, suggesting that the effects of the intervention were sustained postintervention when the community nutrition experts were not receiving feedback on their performance. This is an effect size of Glass's  $\delta=1.690$  (control group SD of duration of counseling for the period is 3.66), and an effect size of using Cohen's  $d=0.813$  (pooled SD of duration of counseling for the period is 5.87), both of which indicated a moderate to large effect of receiving duration of counseling. The results held after correcting for differential attrition, although we had a smaller effect size of using Glass's  $\delta=0.926$  (control group SD of duration of counseling for the period is 4.683) and Cohen's  $d=0.748$  (pooled SD of duration of counseling for the period is 5.795). Because the results after correcting for the attrition bias were similar to the results generated by the original sample, we proceeded with the original sample for further analysis.

**Table 4.** Difference in differences (DID) estimates: impact of calls on performance indicators in the intervention and postintervention stages.

Dependent variable	Case activity			Form submissions (number)			Duration of counseling		
	DID	<i>t</i> (df)	<i>P</i>	DID	<i>t</i> (df)	<i>P</i>	DID	<i>t</i> (df)	<i>P</i>
<b>Original sample (n=500)</b>									
Treatment	-3.98	-0.57 (499)	.57	-7.09	-0.95 (499)	.34	-0.54	-0.32 (499)	.75
After	9.07	2.65 (499)	.008	-32.32	8.20 (499)	.001	0.57	0.77 (499)	.44
Treatment×after	5.58	0.98 (499)	.33	2.032	0.29 (499)	.77	4.78	3.66 (499)	.001
Constant	22.20	5.34 (499)	.001	52.15	12.31 (499)	.001	2.43	2.54 (499)	.01
<b>Correcting for attrition (n=600)</b>									
Treatment	-6.06	-1.40 (599)	.16	-1.142	-0.29 (599)	.77	-0.64	-0.42 (599)	.67
After	12.71	4.52 (599)	.001	-25.36	-7.77 (599)	.001	0.746	1.09 (599)	.27
Treatment×after	4.45	1.2 (599)	.023	-1.39	-0.37 (599)	.71	4.34	3.65 (599)	.001
Constant	21.43	5.18 (599)	.001	44.51	10.7 (599)	.001	2.31	2.62 (599)	.009

We tested for statistical significance of the marginal difference in duration of counseling in the intervention and postintervention periods by restricting the data to only include observations in the postintervention period (weeks 7-10). We ran two specifications, first to look for changes in duration of counseling from the baseline period of week 1 and second where we set the baseline at the end of the intervention period in week 6. At week 1, duration of counseling was mean 2.45 (SD 3.80) minutes for the treatment group (n=34) and mean 1.89 (SD 2.83) minutes for the control group (n=16). At week 6, duration of counseling was mean 3.15 (SD 3.63) minutes for the control group (n=16) and mean 10.70 (SD 6.50) minutes for the treatment group (n=34). At weeks 7 to 10, duration of counseling was mean 3.82 (SD 5.53) minutes for the control group (n=64) and mean 9.22 (SD 7.71) minutes for the treatment group (n=136).

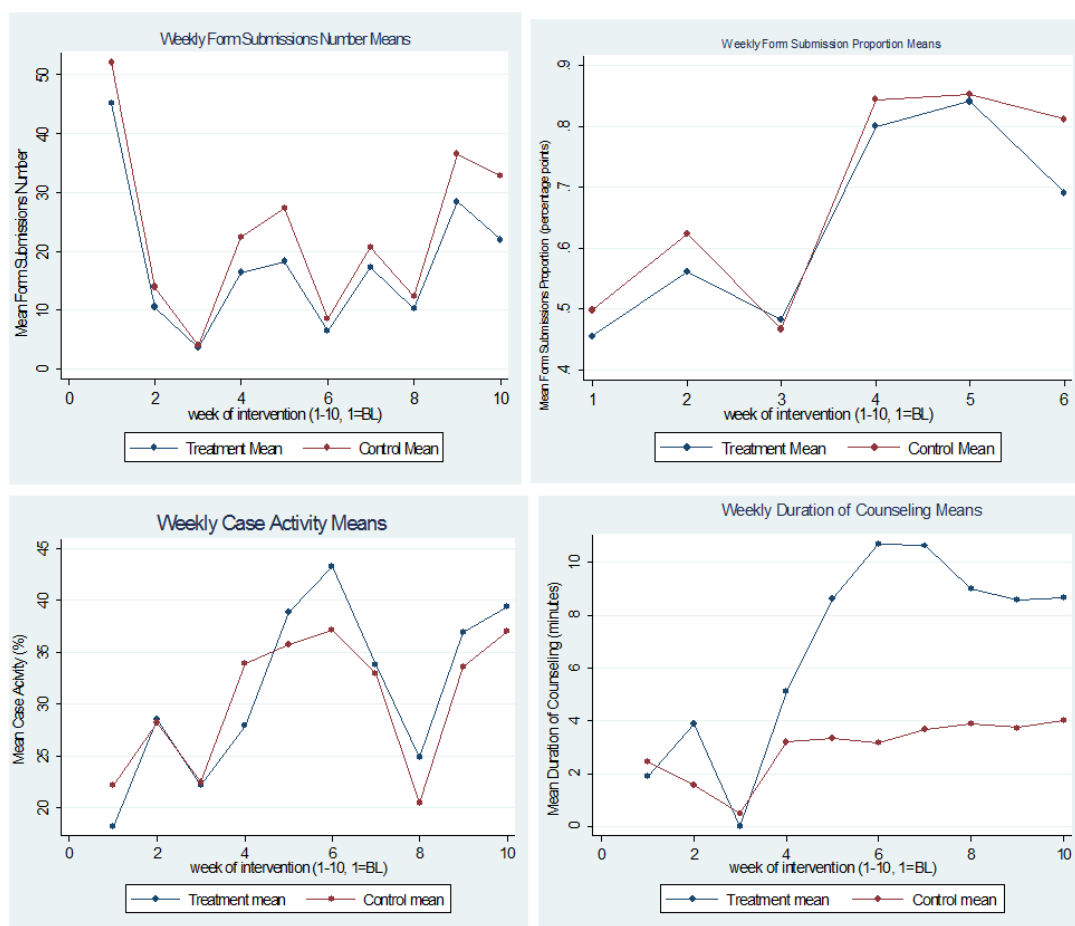
The results of the two specifications are presented in [Table 5](#). We found that there was positive and significant impact of the calls on duration of counseling performance from baseline to the postintervention period using only postintervention data

(coefficient=5.939,  $P<.001$ ). Setting week 6 as the baseline and comparing counseling times at the end of the intervention period to the postintervention period, we found that counseling times fell significantly (coefficient=-2.14,  $P=.02$ ). This result is highlighted in [Figure 4](#), which shows weekly mean counseling times for treatment and control groups over both study periods. This result shows that performance feedback is an important driver for performance where duration of counseling is concerned. Although we found positive intertemporal effects and receiving performance feedback led to large sustained improvements in counseling times for community nutrition experts, they dropped off in the postintervention period when they were not receiving feedback. The community nutrition experts were retrained on using the counseling forms and were reminded of their targets during the calls, the effects of which were sustained through to the postintervention period. However, they performed better while receiving performance feedback and reminders of their targets, and the feedback seemed to have intensified their internal motivation to perform and meet their goals.

**Table 5.** Impact of performance feedback on duration of counseling in the postintervention period looking for sustained effects in performance.

Dependent variable	Duration of counseling					
	Baseline=week 1 (n=250)			Baseline=week 6 (n=250)		
	Coefficient	t (249)	P	Coefficient	t (249)	P
Duration of counseling	-0.536	-0.30	.76	7.552	4.09	.001
After (baseline=week 1)	1.390	2.07	.39			
Counseling×after (baseline=week 1)	5.939	5.00	.001			
After (baseline=week 6)				0.672	1.34	.18
Counseling×after (baseline=week 6)				-2.149	-2.43	.02
Constant	2.433	2.4	.02	3.151	3.02	.002

**Figure 4.** Treatment and control means in intervention and postintervention stages. Duration of counseling means are the mean duration of counseling per form submitted. BL: baseline.



There were no significant effects of receiving feedback on form submissions or case activity, in the intervention and postintervention periods. In the postintervention period, we did not have data for the total number of clients visited in the 7-day period; hence, we could not include form submissions as proportion of clients visited as a variable. Before the intervention, both form submissions and case activity were lower in the treatment group; however, this difference was not significant. There seemed to be an overall decrease in form submissions for all community nutrition experts (after) and the pre-post changes in the treatment and control groups were essentially the same. This could be explained by the fact that

most community nutrition experts had multiple registrations of the same case in their apps, with some community nutrition experts having up to five duplicate registrations of the same case. Before the intervention, the community nutrition experts submitted up to five different infant and nutrition forms for each of the duplicate registrations. This meant that baseline form submissions were much higher than appropriate because the community nutrition experts submitted many duplicate forms for the same case. The performance feedback and retraining given to all community nutrition experts during the study period by the program reinforced that there was no need to submit duplicate forms despite the duplicate registrations, and only one



infant health and nutrition form should be submitted per visit rather than submitting multiple forms for each of the duplicate registrations. This explains the overall decrease in form submissions in the intervention and postintervention period. Our interpretation was supported by the increase in form submissions proportional to number of clients visited in specification 3 for the control group and a slightly lower effect for the treatment group in the postintervention period.

Although case activity also showed an increase in performance in both the intervention and postintervention periods, the results were not significant. However, there was an overall increase in case activity for community nutrition experts in the control group in the intervention and postintervention period. The interpretation here could be that regardless of the information received during the calls, receiving calls boosted the community nutrition experts' case activity. However, because we did not have a control group that did not receive calls, we could not estimate the effect of calling itself. The results are confounded by other factors, which could have caused a secular trend of increasing performance on case activity, including retraining during the intervention period by RMF, which alone could explain the increased performance after the intervention, independent of the intervention itself.

Although there were some program-wide changes to increase community nutrition expert performance during our intervention confounding the effects, including distribution of an additional paper-based job aid for counseling and retraining on CommCare for all the community nutrition experts, our intervention played an important part in increasing community nutrition expert performance for case activity. One interpretation of these results could be that regardless of the information disclosed in the calls, receiving calls boosted community nutrition expert performance by increasing internal motivation and increasing interest in their work. Although providing performance feedback against a target was an important element of our intervention, the feedback calls went beyond providing simple feedback and fluid discussions about the community nutrition experts' technical and program-related challenges, and some personal/family matters were an important part of the calls. Our intervention strengthened supervisory structures already present in the program, regularly communicating the technical and program-related challenges to the project coordinator and the community nutrition experts' immediate supervisors, who were able to resolve these more quickly, contributing to the increased case activity for all community nutrition experts in the postintervention stages. Additionally, the calls also resolved problems the community nutrition experts were having in using their app, which could also have contributed to the increasing their case activity because all data was collected through their CommCare apps.

The challenge in drawing out the impact of the calls is that we did not have a control group that did not receive any calls in our design. However, our interpretation is supported by the qualitative work supporting this study where community nutrition experts mentioned that the calls were an important factor in motivating and engaging them, increasing their interest in their work. The results presented in [Table 3](#) show that the calls were seen as an indication that the program and CommCare

were taking an interest in their work and the community nutrition experts mentioned that it increased their job satisfaction and motivation as a result. In focus group discussions, community nutrition experts mentioned that their interest and motivation grew when they engaged with something new on a periodic basis and had the opportunity to interact with people who were not a part of their usual day-to-day interactions. The calls were able to provide them with such a service, targeting autonomy, relatedness, and competence, and contributing to strengthening their intrinsic motivation and interest toward their work.

The end-line survey administered to all community nutrition experts suggested that community nutrition experts perceived the calls as a motivator having a positive effect on their performance. In all, 86% (38/44) reported that they felt the calls improved their performance, 79% (34/43) felt the calls motivated them a lot to do their work, and 45% (20/44) found performance feedback to be the most effective component of the calls for motivation. In addition, 75% (33/44) reported that their technical problems were resolved faster as a result of the calls and 100% (44/44) of community nutrition experts felt that their knowledge about using CommCare had improved as a result of the calls and 94% (43/46) would sign up for this service. These results support our interpretation that the calls improved community nutrition experts' intrinsic motivation by supporting competence, relatedness, and autonomy, leading to improvements in case activity for all community nutrition experts.

### Cross-Partial Effects

To determine whether receiving feedback on any indicator affected case activity performance or performance on the remaining two indicators, we estimated cross-partial effects of providing performance feedback on one indicator on the performance on the other two indicators. For each of the treatments, the remaining two groups acted as a pooled control and we were interested in estimating cross-partial effects of receiving feedback on case activity, on performance in duration of counseling, or form submissions. We present the analysis for the postintervention period because this generated higher power and was a more policy-relevant result. However, our analysis showed the results in the intervention period were similar to those in the postintervention period. There was one cross-partial effect where feedback on one indicator affected performance on another indicator. It negatively affected duration of counseling by approximately 3.39 minutes ( $P=.01$ ), suggesting that receiving duration of counseling feedback was more effective in increasing counseling times than receiving form submissions feedback. We did not find any cross-partial effects for the remaining two indicators.

### Comparing the Three Groups Head-to-Head

We compared each of the three treatment groups head-to-head and estimated how feedback on each indicator affected each treatment group separately. The reference group in each specification was the treatment group. The results were in-line with our analysis from equation 1 ([Multimedia Appendix 1](#)); there were significant improvements in counseling times with mean counseling times of community nutrition experts in the duration of counseling group increasing compared with those

in the form submissions or case activity treatment groups. Receiving feedback on form submissions or case activity did not significantly impact performance on that indicator compared to the two other groups. The results support the previous

findings, with no baseline differences between the three treatment groups, and duration of counseling feedback affecting counseling times differently than case activity or form submissions feedback (Table 6).

**Table 6.** Head-to-head comparison of treatment groups (n=500).

Variable	Case activity			Form submissions (number)			Duration of counseling		
	Coefficient	t (499)	P	Coefficient	t (499)	P	Coefficient	t (499)	P
<b>Treatment</b>									
Case activity	—	—	—	1.99	0.23	.81	0.19	0.09	.92
Form submissions	3.66	0.46	.64	—	—	—	0.94	0.47	.63
Duration of counseling	4.29	0.54	.58	12.81	1.46	.14	—	—	—
After	4.52	0.93	.35	-3.29	-5.28	.001	5.35	4.97	.001
<b>Treatment×after</b>									
Case activity	10.13	1.53	.12	2.36	0.3	.76	-4.09	-2.77	.005
Duration of counseling	9.10	1.33	.18	-6.97	-0.86	.39			
Form submission							-5.56	-3.65	.001
Constant	18.23	3.31		45.06	7.28		1.89	1.35	.17

### Heterogeneous Effects

The distribution of calls received among the 50 community nutrition experts who were included in the analysis was as follows: 10% (5/50) of community nutrition experts received zero calls, 6% (3/50) received one call, 6% (3/50) received two calls, 16% (8/50) received three calls, 14% (7/50) received four calls, 10% (5/50) received five calls, and 38% (19/50) received all six calls.

We wanted to examine whether the effect of our intervention varied across the intensity of treatment (estimate equation 4 in Multimedia Appendix 1). We interacted the number of calls

received with the treatment and time period (intervention and postintervention) to look for heterogeneous impacts based on intensity of exposure to treatment. Table 7 presents the heterogeneous impacts of call intensity for the postintervention period. The estimates indicated that call intensity mattered for sustaining increases in duration of counseling and counseling times increased by approximately 1.366 minutes ( $P=.04$ ) with each additional call in the postintervention period. The estimates did not indicate any heterogeneous impacts of call intensity on the other two indicators in the intervention or the postintervention periods. The results held after accounting for differential attrition.

**Table 7.** Heterogeneous effects: the effect of receiving more calls on performance.

Variable	Case activity			Form submissions, number			Duration of counseling		
	Coefficient	<i>t</i> ( <i>df</i> )	<i>P</i>	Coefficient	<i>t</i> ( <i>df</i> )	<i>P</i>	Coefficient	<i>t</i> ( <i>df</i> )	<i>P</i>
<b>Original sample (n=500)</b>									
Treatment	-4.70	0.28 (499)	.77	8.05	0.52 (499)	.60	-1.45	-0.40 (499)	.69
After	22.07	3.15 (499)	.001	0.21	0.02 (499)	.98	-0.72	-0.45 (499)	.65
Calls	4.52	2.26 (499)	.02	7.98	3.70 (499)	.001	-0.39	-0.89 (499)	.37
Treatment×after	13.98	1.03 (499)	.30	-9.97	-0.70 (499)	.48	-0.45	-0.16 (499)	.87
Treatment×calls	-0.54	-0.15 (499)	.88	-2.96	-0.84 (499)	.40	0.22	0.27 (499)	.79
After×calls	-3.47	-2.12 (499)	.03	-7.79	-3.93 (499)	.001	0.32	0.90 (499)	.37
Treatment×after×calls	-1.35	-0.47 (499)	.63	2.12	0.65 (499)	.51	1.37	2.11 (499)	.04
Constant	5.24	0.61 (499)	.54	18.81	1.89 (499)	.06	4.05	1.99 (499)	.05
<b>Correcting for attrition (n=600)</b>									
Treatment	-6.28	-1.35 (599)	.17	-0.39	-0.10 (599)	.92	-1.18	-0.38 (599)	.70
After	26.17	4.72 (599)	.001	4.31	0.64 (599)	.52	-0.15	-0.11 (599)	.91
Calls	4.28	2.75 (599)	.006	8.04	5.08 (599)	.001	-0.26	-0.65 (599)	.52
Treatment×after	10.86	1.40 (599)	.16	-0.77	-0.11 (599)	.91	-0.90	-0.36 (599)	.72
Treatment×calls	-0.18	-0.12 (599)	.90	-0.64	-0.39 (599)	.69	0.14	0.19 (599)	.84
After×calls	-3.67	-2.67 (599)	.007	-7.93	-4.97 (599)	.001	0.24	0.75 (599)	.45
Treatment×after×calls	-1.26	-0.64 (599)	.52	-0.02	-0.01 (599)	.99	1.42	2.43 (599)	.02
Constant	5.71	0.83 (599)	.40	14.44	2.06 (599)	.39	3.28	1.91 (599)	.06

The number of calls received per community nutrition expert suggested that the number of calls received, regardless of the indicator for feedback, had a positive impact in pre-post differences in the control group. Because all community nutrition experts received calls, this finding suggested that regardless of the information disclosed in the calls, receiving calls improved performance by providing supportive supervision and increasing intrinsic motivation by including appreciation, encouragement, avenues to discuss problems and solutions, and solving technical difficulties.

## Discussion

Our findings show some interesting differences in the effect our intervention has on the different performance indicators. We found positive impacts on duration of counseling sustained in the postintervention period, whereas case activity and form submissions showed no significant effects. We also found heterogeneous impacts of call intensity for duration of counseling, with each additional call improving the community nutrition expert's counseling times further. There was an overall increase in case activity for all community nutrition experts, whereas the number of form submissions decreased due to

multiple registrations and duplicate form submissions declining indicating improvement in overall reporting standards. We discuss possible reasons for these differing results subsequently, further supplemented with qualitative interviews with community nutrition experts to better understand the results.

First, goal-setting theory suggests that the difficulty of the target matters for motivation and performance, and moderately difficult targets induce more effort and motivation compared to difficult or easier targets [27]. In our study, targets for duration of counseling were easier to achieve because they could be achieved by spending more time with the clients that the community nutrition expert was visiting. For case activity, the target was much more difficult, especially if the community nutrition expert had a lot of cases that she had to follow up on. The target for form submissions was related to case activity and should be an easy one. However, we saw a sharp decline in form submissions for all community nutrition experts because they were asked to stop submitting duplicate registrations and infant health and nutrition forms for a single client.

Second, the data quality for duration of counseling is likely to include less measurement error compared with data capturing case activity and form submissions. The community nutrition experts had many duplicate registrations in their app for each client due to mistraining issues at the beginning of the CommCare program. As a result, some community nutrition experts had up to 400 or 500 clients registered in their app with most of these being duplicate registrations. Case activity was measured against total clients and community nutrition experts were regularly submitting multiple forms (for each of the duplicate registrations) skewing form submissions and case activity data. The overall decrease in form submissions number in both the treatment and control groups can be explained by the retraining from RMF, which reinforced that the community nutrition experts should not submit multiple forms for the same visit in spite of the duplicate registrations. Our intervention also included training on how to correctly use the app and reinforced the appropriate workflow that the community nutrition experts should adopt, which also contributed to the reduction in the number of form submissions. Form submission as a proportion of clients visited did not show any significant decrease supporting our reasoning.

Third, case activity was affected by the holiday period of Diwali, which fell around week 3 of the intervention period. Figure 4 shows the weekly means for case activity for the treatment and control groups, both of which were close to zero during the holiday period. Duration of counseling was not affected by the holiday period because if the community nutrition expert did not visit and counsel a family, it did not affect the mean counseling time.

Fourth, all community nutrition experts also received a paper-based counseling aid during the study, which could also drive the increase in duration of counseling. However, we expect this aid to affect counseling times for all community nutrition experts and not just those in the treatment group. It could also be the case that the additional counseling aid interacted with receiving duration of counseling feedback by positively affecting duration of counseling and, in the absence of the additional

counseling aid, there would be no significant effects on duration of counseling. However, we cannot perform this analysis due to data constraints.

Lastly, the significance of differential attrition by rank is higher in the duration of counseling treatment group than in the other two groups. Thus, it could be the case that the duration of counseling results are being driven by differential attrition, although this is unlikely because all groups show significant differential attrition by rank.

We expect that the changes in performance also include improvements in reporting standards and are being driven by increased effort on the part of the community nutrition expert, where our intervention improved community nutrition expert motivation and encouraged and increased her interest toward her work. External motivation played a part where community nutrition experts are weary of the consequences of bad performance, such as getting fired or reprimanded by the program. Internalization of the programs' goals to improve intrinsic motivation was facilitated by improving the community nutrition experts' competence; the community nutrition experts are better at using their CommCare app after the intervention, which included components where community nutrition experts were trained on proper use of the app to support their work flow. Improvements in duration of counseling are also driven by improvements in reporting standards and increased knowledge in correct use of the app. Before the intervention, many community nutrition experts counseled their clients without opening the family counseling form leading to very low counseling times recorded on CommCare. After the intervention, and retraining on how best to use the family counseling form (ie, we encouraged them to open the form and fill in the details as they counsel), we see an increase in counseling time as a direct impact of our intervention. Accurate reporting is an essential component of performance in the CHW context due to the remote nature of their work and lack of immediate supervision, which means that any CHW's activities, services provided, and changes in the client's health status are not known without accurate data provided by the CHW. We also see improvements in reporting standards for form submissions.

Qualitative interviews with three community nutrition experts after the postintervention period support our interpretation that the calls improved the community nutrition experts' competence and autonomy by improving knowledge and skills and helping to quickly solve any technical issues faced by the community nutrition expert. The continuous monitoring provided by the calls was seen to impact duration of counseling more than the other indicators, with the community nutrition expert giving more time to the client in addition to using the family counseling form while counseling to provide accurate reporting on this indicator.

Case activity also shows overall improvements for all community nutrition experts after our intervention, which is sustained in the postintervention period. Although we do not have a control group that did not receive any calls to provide concrete evidence, the results suggest that the calls, which included a strong component of supportive supervision, helped support autonomy and the internalization of program goals

(which all community nutrition experts were aware of) leading to more autonomous motivation in the community nutrition experts and higher performance across all treatment groups. Our hypothesis that the calls lead to greater autonomous motivation is supported by the sustained improvements in duration of counseling in the postintervention period as a result of receiving feedback calls on duration of counseling and overall improvements in case activity, and improvements in reporting form submissions sustained after the intervention as well.

Providing CHWs with regular performance feedback through calls placed by a supervisor or manager is an effective way to reinforce goals and targets, and provide supportive supervision to CHWs encouraging their extrinsic and autonomous motivation and performance. Two-way communication opens avenues to discuss and provide solutions addressing gaps in performance, any challenges faced at work, or technical or systematic issues afflicting the CHW preventing her/him from performing effectively. Strategies targeting improvements in CHW programs should include an element of continuous monitoring and feedback system reinforced with supportive supervision to generate improvements in CHW performance and maintain an effective CHW program.

Other than the limitations posed by the dataset, our study has some important limitations that could potentially bias our outcomes:

1. Risks of unethical behavior leading to false reporting.
2. Spillovers where targets for different indicators are known to community nutrition experts in other groups.
3. Low network connectivity or unwillingness/inability of the community nutrition experts to respond to and answer the calls (could potentially be endogenous to performance if the community nutrition experts are aware of their own performance and do not answer the call if they perceive their performance to be low).

4. Holidays and events in week 3, which meant that the community nutrition expert(s) were absent with consent from their managers. Some may have extended their break. Attendance is a prerequisite to performance.

5. The most important limitation in the study arises because we do not have a control group that received no calls and no performance feedback. Hence, we cannot discern the effects of the calls themselves. This is important because the calls included other components other than the performance feedback, which can impact the community nutrition experts' performance across all indicators, including the indicators on which they received no performance feedback.

In order to limit and test for the bias, we adopted the following strategies:

1. Draw an upper limit on the indicators where false reporting is most plausible. We imposed a limit of 20 minutes on the duration of counseling so that any outliers did not skew the mean counseling times.
2. Although we ensured that no community nutrition expert received feedback on other indicators other than the treatment she was assigned to, RMF reiterated the targets for all three indicators during the course of the study, which could have affected the community nutrition experts' performance by setting goals. Goal setting can motivate performance, but we were interested in the effects of goal setting in conjunction with providing feedback.
3. We tested for correlation between initial performance and number of calls received to test whether community nutrition experts were deliberately avoiding calls anticipating unfavorable performance feedback. The pairwise correlation for baseline rank and number of calls received was  $r=-.265$  ( $P<.001$ ) suggesting that lower-performing community nutrition experts were less likely to answer calls.

---

## Acknowledgments

We would like to thank USAID for funding the research, as well as RMF for their participation in the study.

---

## Conflicts of Interest

Dimagi is the owner of CommCare, the mHealth platform used by the participants in the study. Authors Neal Lesh and Sangya Kaphle were employed by Dimagi during the time of research. Michael-Mathke Ficher works with RMF.

---

## Multimedia Appendix 1

Equations used to estimate main treatment effects and heterogeneous effects.

[[PDF File \(Adobe PDF File\), 53KB - publichealth\\_v2i2e169\\_app1.pdf](#)]

---

## Multimedia Appendix 2

Results of matching based on ranks and treatment groups.

[[PDF File \(Adobe PDF File\), 37KB - publichealth\\_v2i2e169\\_app2.pdf](#)]

---

## References



1. Prasad BM, Muraleedharan V. Community Health Workers: a Review of Concepts, Practice and Policy Concerns. 2008 Mar. London: UK Department for International Development. URL: [http://www.crehs.lshtm.ac.uk/Community\\_health\\_workers\\_prasad.pdf](http://www.crehs.lshtm.ac.uk/Community_health_workers_prasad.pdf) [accessed 2014-03-06] [WebCite Cache ID 6NriMAGoS]
2. Pérez LM, Martínez J. Community health workers: social justice and policy advocates for community health and well-being. *Am J Public Health* 2008 Jan;98(1):11-14. [doi: [10.2105/AJPH.2006.100842](https://doi.org/10.2105/AJPH.2006.100842)] [Medline: [18048789](https://pubmed.ncbi.nlm.nih.gov/18048789/)]
3. Farmer P, Léandre F, Mukherjee JS, Claude M, Nevil P, Smith-Fawzi MC, et al. Community-based approaches to HIV treatment in resource-poor settings. *Lancet* 2001 Aug 4;358(9279):404-409. [Medline: [11502340](https://pubmed.ncbi.nlm.nih.gov/11502340/)]
4. Patel AR, Nowalk MP. Expanding immunization coverage in rural India: a review of evidence for the role of community health workers. *Vaccine* 2010 Jan 8;28(3):604-613. [doi: [10.1016/j.vaccine.2009.10.108](https://doi.org/10.1016/j.vaccine.2009.10.108)] [Medline: [19896450](https://pubmed.ncbi.nlm.nih.gov/19896450/)]
5. Swider SM. Outcome effectiveness of community health workers: an integrative literature review. *Public Health Nurs* 2002;19(1):11-20. [Medline: [11841678](https://pubmed.ncbi.nlm.nih.gov/11841678/)]
6. UNICEF. What Works for Children in South Asia: Community Health Workers. Kathmandu, Nepal: The United Nations Children's Fund (UNICEF); 2004. URL: <http://www.unicef.org/rosa/community.pdf> [accessed 2014-03-06] [WebCite Cache ID 6Nribs8bO]
7. Franco LM, Bennett S, Kanfer R. Health sector reform and public sector health worker motivation: a conceptual framework. *Soc Sci Med* 2002 Apr;54(8):1255-1266. [Medline: [11989961](https://pubmed.ncbi.nlm.nih.gov/11989961/)]
8. Rowe AK, de Savigny D, Lanata CF, Victora CG. How can we achieve and maintain high-quality performance of health workers in low-resource settings? *Lancet* 2005;366(9490):1026-1035. [doi: [10.1016/S0140-6736\(05\)67028-6](https://doi.org/10.1016/S0140-6736(05)67028-6)] [Medline: [16168785](https://pubmed.ncbi.nlm.nih.gov/16168785/)]
9. Dieleman M, Cuong PV, Anh LV, Martineau T. Identifying factors for job motivation of rural health workers in North Viet Nam. *Hum Resour Health* 2003 Nov 5;1(1):10 [FREE Full text] [doi: [10.1186/1478-4491-1-10](https://doi.org/10.1186/1478-4491-1-10)] [Medline: [14613527](https://pubmed.ncbi.nlm.nih.gov/14613527/)]
10. Gopalan SS, Mohanty S, Das A. Assessing community health workers' performance motivation: a mixed-methods approach on India's Accredited Social Health Activists (ASHA) programme. *BMJ Open* 2012;2(5) [FREE Full text] [doi: [10.1136/bmjopen-2012-001557](https://doi.org/10.1136/bmjopen-2012-001557)] [Medline: [23019208](https://pubmed.ncbi.nlm.nih.gov/23019208/)]
11. Mbindyo P, Gilson L, Blaauw D, English M. Contextual influences on health worker motivation in district hospitals in Kenya. *Implement Sci* 2009;4:43 [FREE Full text] [doi: [10.1186/1748-5908-4-43](https://doi.org/10.1186/1748-5908-4-43)] [Medline: [19627590](https://pubmed.ncbi.nlm.nih.gov/19627590/)]
12. Kok MC, Muula AS. Motivation and job satisfaction of health surveillance assistants in Mwanza, Malawi: an explorative study. *Malawi Med J* 2013 Mar;25(1):5-11 [FREE Full text] [Medline: [23717748](https://pubmed.ncbi.nlm.nih.gov/23717748/)]
13. Ghimire J, Gupta RP, Kumal AB, Mahato RK, Bhandari RM, Thapa N. Factors associated with the motivation and de-motivation of health workforce in Nepal. *J Nepal Health Res Counc* 2013 May;11(24):112-118. [Medline: [24362597](https://pubmed.ncbi.nlm.nih.gov/24362597/)]
14. Gagné M, Deci E. Self-determination theory and work motivation. *J Organiz Behav* 2005 Apr 14;26(4):331-362 [FREE Full text] [doi: [10.1002/job.322](https://doi.org/10.1002/job.322)]
15. Shanks N, Dore A. Management and motivation. In: Buchbinder SB, Shanks NH, editors. *Introduction to Health Care Management*. 2nd Edition. London: Jones and Bartlett; 2012:39-52.
16. Ryan RM, Deci EL. Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *Am Psychol* 2000 Jan;55(1):68-78. [Medline: [11392867](https://pubmed.ncbi.nlm.nih.gov/11392867/)]
17. Deci EL, Koestner R, Ryan RM. A meta-analytic review of experiments examining the effects of extrinsic rewards on intrinsic motivation. *Psychol Bull* 1999 Nov;125(6):627-68; discussion 692. [Medline: [10589297](https://pubmed.ncbi.nlm.nih.gov/10589297/)]
18. Deci EL. Effects of externally mediated rewards on intrinsic motivation. *J Pers Soc Psychol* 1971 Dec 01;18(1):105-115.
19. Erez M, Gopher D, Arzi N. Effects of goal difficulty, self-set goals, and monetary rewards on dual task performance. *Organ Behav Hum Dec* 1990 Dec;47(2):247-269. [doi: [10.1016/0749-5978\(90\)90038-B](https://doi.org/10.1016/0749-5978(90)90038-B)]
20. Ryan RM, Mims V, Koestner R. Relation of reward contingency and interpersonal context to intrinsic motivation: a review and test using cognitive evaluation theory. *J Pers Soc Psychol* 1983;45(4):736-750. [doi: [10.1037/0022-3514.45.4.736](https://doi.org/10.1037/0022-3514.45.4.736)]
21. Deci E, Ryan R. *Intrinsic Motivation and Self-Determination in Human Behavior (Perspectives in Social Psychology)*. New York: Plenum Publishing Co; 1985.
22. Joussemet M, Koestner R, Lekes N, Houliort N. Introducing uninteresting tasks to children: a comparison of the effects of rewards and autonomy support. *J Personality* 2004 Feb;72(1):139-166. [doi: [10.1111/j.0022-3506.2004.00259.x](https://doi.org/10.1111/j.0022-3506.2004.00259.x)]
23. Baard PP, Deci EL, Ryan RM, Houliort N. The relation of intrinsic need satisfaction to performance and wellbeing in two work settings. *J Appl Psychol* 2004;34(10):2045-2068.
24. Deci EL, Eghrari H, Patrick BC, Leone DR. Facilitating internalization: the self-determination theory perspective. *J Personality* 1994 Mar;62(1):119-142. [doi: [10.1111/j.1467-6494.1994.tb00797.x](https://doi.org/10.1111/j.1467-6494.1994.tb00797.x)]
25. Baard P, Deci E, Ryan R. Intrinsic need satisfaction: a motivational basis of performance and well-being in two work settings. *J Appl Social Psychol* 2004 Oct;34(10):2045-2068. [doi: [10.1111/j.1559-1816.2004.tb02690.x](https://doi.org/10.1111/j.1559-1816.2004.tb02690.x)]
26. Locke E, Latham G, Smith K, Wood R. *A Theory of Goal Setting and Task Performance*. Englewood Cliffs, NJ: Prentice-Hall College Division; 1990.
27. Locke EA, Latham GP. Building a practically useful theory of goal setting and task motivation. A 35-year odyssey. *Am Psychol* 2002 Sep;57(9):705-717. [Medline: [12237980](https://pubmed.ncbi.nlm.nih.gov/12237980/)]
28. Matsui T, Okada A, Kakuyama T. Influence of achievement need on goal setting, performance, and feedback effectiveness. *J Appl Psychol* 1982;67(5):645-648. [doi: [10.1037/0021-9010.67.5.645](https://doi.org/10.1037/0021-9010.67.5.645)]

29. Bandura A, Cervone D. Differential engagement of self-reactive influences in cognitive motivation. *Organ Behav Hum Dec* 1986 Aug;38(1):92-113. [doi: [10.1016/0749-5978\(86\)90028-2](https://doi.org/10.1016/0749-5978(86)90028-2)]
30. Becker LJ. Joint effect of feedback and goal setting on performance: a field study of residential energy conservation. *J Appl Psychol* 1978;63(4):428-433. [doi: [10.1037/0021-9010.63.4.428](https://doi.org/10.1037/0021-9010.63.4.428)]
31. Erez M. Feedback: a necessary condition for the goal-performance relationship. *J Appl Psychol* 1977;62(5):624-627.
32. Strang H, Lawrence E, Fowler P. Effects of assigned goal level and knowledge of results on arithmetic compensation: A laboratory study. *J Appl Psychol* 1978 May 01;63:446.
33. Ordonez LD, Schweitzer ME, Galinsky AD, Bazerman MH. Goals Gone Wild: The Systematic Side Effects of Overprescribing Goal Setting. *Acad Manage Perspect* 2009 Feb 01;23(1):6-16. [doi: [10.5465/AMP.2009.37007999](https://doi.org/10.5465/AMP.2009.37007999)]
34. Dickin KL, Dollahite JS, Habicht J. Enhancing the intrinsic work motivation of community nutrition educators: how supportive supervision and job design foster autonomy. *J Ambul Care Manage* 2011;34(3):260-273. [doi: [10.1097/JAC.0b013e31821dc63b](https://doi.org/10.1097/JAC.0b013e31821dc63b)] [Medline: [21673524](https://pubmed.ncbi.nlm.nih.gov/21673524/)]
35. Bajpai N, Dholakia R. Working Paper No 1: Improving the Performance of Accredited Social Health Activists in India. Nariman Point, Mumbai: Columbia Global Centers; 2011 May. URL: [http://globalcenters.columbia.edu/files/cgc/pictures/Improving\\_the\\_Performance\\_of\\_ASHAs\\_in\\_India\\_CGCSA\\_Working\\_Paper\\_1.pdf](http://globalcenters.columbia.edu/files/cgc/pictures/Improving_the_Performance_of_ASHAs_in_India_CGCSA_Working_Paper_1.pdf) [accessed 2014-03-06] [WebCite Cache ID 6NrjojjvE]
36. Robinson SA, Larsen DE. The relative influence of the community and the health system on work performance: A case study of community health workers in Colombia. *Soc Sci Med* 1990 Jan;30(10):1041-1048. [doi: [10.1016/0277-9536\(90\)90290-9](https://doi.org/10.1016/0277-9536(90)90290-9)]
37. Bhattacharyya K, Winch P, LeBan K, Tien M. Community Health Worker Incentives and Disincentives: How They Affect Motivation, Retention, and Sustainability. Arlington, VA: Basic Support for Institutionalizing Child Survival Project (BASICS II) for the United States Agency for International Development; 2001.
38. Bhattacharji S, Abraham S, Muliyl J, Job JS, John K, Joseph A. Evaluating community health worker performance in India. *Health Policy Plan* 1986;1(3):232-239. [doi: [10.1093/heapol/1.3.232](https://doi.org/10.1093/heapol/1.3.232)]
39. Frimpong JA, HELLERINGER S, Awoonor-Williams JK, Yeji F, Phillips JF. Does supervision improve health worker productivity? Evidence from the Upper East Region of Ghana. *Trop Med Int Health* 2011 Oct;16(10):1225-1233 [FREE Full text] [doi: [10.1111/j.1365-3156.2011.02824.x](https://doi.org/10.1111/j.1365-3156.2011.02824.x)] [Medline: [21729221](https://pubmed.ncbi.nlm.nih.gov/21729221/)]
40. Treatman D, Lesh N. Strengthening community health systems with localized multimedia. In: Proceedings of M4D 2012. 2012 May 01 Presented at: M4D 2012; Feb 28-29, 2012; New Delhi, India p. 7-22 URL: <http://mobileeducationkit.net/doc/Mek7-FULLTEXT01.pdf#page=24>
41. DeRenzi B, Findlater L, Payne J, Birnbaum B, Mangilima J, Parikh T, et al. Improving community health worker performance through automated SMS. In: Proceedings of the Fifth International Conference on Information and Communication Technologies and Development. 2012 Presented at: ICTD '12 Fifth International Conference on Information and Communication Technologies and Development; Atlanta, GA; Mar 12-15, 2012 p. 25-34. [doi: [10.1145/2160673.2160677](https://doi.org/10.1145/2160673.2160677)]
42. Cameron J, Pierce WD. Reinforcement, reward, and intrinsic motivation: a meta-analysis. *Rev Educ Res* 1994 Jan 01;64(3):363-423. [doi: [10.3102/00346543064003363](https://doi.org/10.3102/00346543064003363)]
43. Eisenberger R, Cameron J. Detrimental effects of reward. Reality or myth? *Am Psychol* 1996 Nov;51(11):1153-1166. [Medline: [8937264](https://pubmed.ncbi.nlm.nih.gov/8937264/)]
44. Mathauer I, Imhoff I. Health worker motivation in Africa: the role of non-financial incentives and human resource management tools. *Hum Resour Health* 2006;4:24 [FREE Full text] [doi: [10.1186/1478-4491-4-24](https://doi.org/10.1186/1478-4491-4-24)] [Medline: [16939644](https://pubmed.ncbi.nlm.nih.gov/16939644/)]
45. Gneezy U, Rustichini A. Pay enough or don't pay at all. *Q J Econ* 2000 Aug;115(3):791-810. [doi: [10.1162/003355300554917](https://doi.org/10.1162/003355300554917)]

## Abbreviations

- CHW:** community health worker
- DID:** difference in difference
- RMF:** Real Medicine Foundation
- SDT:** self-determination theory

*Edited by G Eysenbach; submitted 06.03.14; peer-reviewed by B Theis, S Panday; comments to author 15.07.14; revised version received 14.03.15; accepted 14.03.16; published 07.12.16.*

*Please cite as:*

*Kaphle S, Matheke-Fischer M, Lesh N*

*Effect of Performance Feedback on Community Health Workers' Motivation and Performance in Madhya Pradesh, India: A Randomized Controlled Trial*

*JMIR Public Health Surveill 2016;2(2):e169*

URL: <http://publichealth.jmir.org/2016/2/e169/>

doi: [10.2196/publichealth.3381](https://doi.org/10.2196/publichealth.3381)

PMID: [27927607](https://pubmed.ncbi.nlm.nih.gov/27927607/)

©Sangya Kaphle, Michael Matheke-Fischer, Neal Lesh. Originally published in JMIR Public Health and Surveillance (<http://publichealth.jmir.org>), 07.12.2016. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Public Health and Surveillance, is properly cited. The complete bibliographic information, a link to the original publication on <http://publichealth.jmir.org>, as well as this copyright and license information must be included.

---

Publisher:  
JMIR Publications  
130 Queens Quay East.  
Toronto, ON, M5A 3Y5  
Phone: (+1) 416-583-2040  
Email: [support@jmir.org](mailto:support@jmir.org)

---

<https://www.jmirpublications.com/>