Journal of Medical Internet Research

Contents

Reviews

Original Papers

Review

Technologies That Assess the Location of Physical Activity and Sedentary Behavior: A Systematic Review

Adam Loveday[1,2], MSc; Lauren B Sherar[1,2], PhD; James P Sanders[1,2], MSc; Paul W Sanderson[1], PhD; Dale W Esliger[1,2], PhD

[1]National Centre for Sport and Exercise Medicine, School of Sport, Exercise and Health Sciences, Loughborough University, Loughborough, United Kingdom

[2]National Institute for Health Research (NIHR) Leicester-Loughborough Diet, Lifestyle and Physical Activity Bio-medical Research Unit, Leicester, United Kingdom

Corresponding Author:
Adam Loveday, MSc
National Centre for Sport and Exercise Medicine
School of Sport, Exercise and Health Sciences
Loughborough University
Epinal Way
Loughborough, LE11 3TU
United Kingdom
Phone: 44 7920015099
Fax: 44 1509 226310
Email: a.loveday@lboro.ac.uk

Abstract

Background: The location in which physical activity and sedentary behavior are performed can provide valuable behavioral information, both in isolation and synergistically with other areas of physical activity and sedentary behavior research. Global positioning systems (GPS) have been used in physical activity research to identify outdoor location; however, while GPS can receive signals in certain indoor environments, it is not able to provide room- or subroom-level location. On average, adults spend a high proportion of their time indoors. A measure of indoor location would, therefore, provide valuable behavioral information.

Objective: This systematic review sought to identify and critique technology which has been or could be used to assess the location of physical activity and sedentary behavior.

Methods: To identify published research papers, four electronic databases were searched using key terms built around behavior, technology, and location. To be eligible for inclusion, papers were required to be published in English and describe a wearable or portable technology or device capable of measuring location. Searches were performed up to February 4, 2015. This was supplemented by backward and forward reference searching. In an attempt to include novel devices which may not yet have made their way into the published research, searches were also performed using three Internet search engines. Specialized software was used to download search results and thus mitigate the potential pitfalls of changing search algorithms.

Results: A total of 188 research papers met the inclusion criteria. Global positioning systems were the most widely used location technology in the published research, followed by wearable cameras, and radio-frequency identification. Internet search engines identified 81 global positioning systems, 35 real-time locating systems, and 21 wearable cameras. Real-time locating systems determine the indoor location of a wearable tag via the known location of reference nodes. Although the type of reference node and location determination method varies between manufacturers, Wi-Fi appears to be the most popular method.

Conclusions: The addition of location information to existing measures of physical activity and sedentary behavior will provide important behavioral information.

XSL•FO
RenderX

Introduction

Physical activity has a long-established relationship with several chronic conditions including diabetes, heart disease, and certain forms of cancer [1]. Recent evidence suggests that sedentary behavior carries deleterious effects on health outcomes independent of moderate-to-vigorous physical activity (MVPA) in young people [2] and adults [3], although this is not a uniform finding [4]. Sedentary behaviors are defined as any waking activity with an energy expenditure of ≤1.5 metabolic equivalents (METs) while in a sitting or reclining position [5]. A paradigm shift is underway toward an increasing appreciation of the importance of reducing sedentary time alongside increasing physical activity [6].

Within the behavioral epidemiology framework [7], the location of a behavior may influence the correlates of the behavior and the intervention strategies needed to change behavior. Discerning the varying contribution of multiple locations to physical activity and sedentary time will allow researchers to target interventions to locations which are associated with the lowest levels of physical activity or highest levels of sedentary time. Understanding the contribution of multiple locations to health behaviors first requires the accurate measurement of location, as suggested by the behavioral epidemiology framework [7,8].

Sedentary behavior and physical activity differ in the domains and locations in which they are likely to occur. Sedentary time is likely, though not exclusively, to occur indoors at the home, at work or school, or in leisure pursuits such as eating a meal or going to the cinema. Conversely, MVPA may occur through active transport, housework, or purposeful exercise. This can be illustrated through the close link between adults, on average, spending approximately 90% of time indoors [9,10] and approximately 60% of time in sedentary activities [11]. The large proportion of time spent indoors and the increasing research focus on sedentary behavior suggest that an accurate measure of where behavior occurs indoors would be particularly valuable.

Determining where physical activity and sedentary time are performed will provide valuable information in isolation; however, it can also act in a synergistic manner. For example, much recent effort has focused on the use of complex pattern recognition techniques to determine the mode or type of activity being performed from raw acceleration data. Depending on the classification method used, classification accuracies between 50% and 90% have been achieved [12]. Given the probabilistic nature of these activity classification methods, the inclusion of location-based data into the current algorithms may provide greater levels of accuracy. For instance, the likelihood of stair climbing is greatly increased if an individual is near a staircase. Similarly, context-sensitive questioning via ecological momentary assessment (EMA) [13] can be enhanced by using location to trigger desirable questions in place of time-based cues.

Furthermore, measurement of indoor location could benefit research into the correlates of physical activity or sedentary behavior. For example, the presence of a television set in a child's bedroom may be a correlate of higher screen time [14]; however, this may be a stronger correlate for those who spend more time in their bedrooms. Establishing how much time a child spends in their bedroom via objective indoor location could, therefore, fully elucidate the strength of this correlate. Thus, the accurate measurement of location could greatly enhance several areas within physical activity and sedentary behavior research, both in and of itself and as an adjunct to other research areas.

Individuals may be able to accurately report the broad location of their physical activity and sedentary behavior [15]; however, self-report location instruments are unable to provide detailed and temporally patterned location information. Objective monitoring could, therefore, provide a more robust means to measure the location of physical activity and sedentary behavior. To date, time indoors has been inferred through the lack of a global positioning system (GPS) signal [16] or through the use of a light (lux) sensor incorporated into activity monitors [17]. However, these methods are only able to differentiate indoor from outdoor and do not provide room- or subroom-level location. Alongside measures of outdoor location, there is, therefore, a need for measures of room- and subroom-level indoor locations, which are feasible for use in this field of research. This review aims to provide an overview of devices and technology currently used, or that could potentially be used, to assess the indoor or outdoor location of physical activity and/or sedentary behavior.

Methods

Search Strategy

Search strategies to identify potentially relevant articles were built around three key groups of keywords: behavior, measurement, and context. Key terms were as follows: sedentary lifestyle, sedentary lifestyles, sedentary behav*, screen time, seden*, sitting time, motor activity, motor activities, physical activity, or activities of daily living; measur*, assess*, patterns, monitor, or sensor; and context*, setting, location, mode, domains, or environment. Scopus, Web of Science, PubMed, Institute of Electrical and Electronics Engineers (IEEE), and OpenGrey were searched using the key terms up to February 4, 2015. Subsequently, forward and backward searching of included articles (ie, references and articles citing the included article) was conducted to identify any further eligible articles. In addition, manual searches of personal files were conducted.

Inclusion and Exclusion Criteria

To be included in this review, studies were required to meet the following criteria: (1) be published in the English language, (2) either describe a tool used to measure the location of physical activity and/or sedentary behavior or provide sufficient information to discern whether the instrument could be modified to measure location, and (3) be a portable/wearable tool. Technologies were required to be portable or wearable to ensure that the technology is always with the participant and that the scope of the review was not so broad as to be unmanageable by including nonwearable technologies (eg, closed-circuit television [CCTV]). A minimum of one part of the measurement system, not the whole system, was required to be wearable/portable for

inclusion. For example, GPS systems consist of a wearable unit and orbiting satellites (ie, one part of the system is wearable but the whole system also consists of unwearable components). Wearable technologies is also an area which is experiencing rapid growth in the consumer sector, as technology increasingly becomes smaller, more powerful, and multi-purpose. Wearable technologies, therefore, give this review a contemporary positioning. No date restriction was placed on search results. Studies erroneously defining sedentary behavior as the absence of sufficient physical activity rather than activities undertaken in a sitting or reclined position [5], were treated as physical activity studies.

Identification of Relevant Studies

Titles and then abstracts of identified articles were screened to determine eligibility based on the above inclusion criteria. Titles and abstracts which did not meet the inclusion criteria were excluded. Following this, the full text of any potentially relevant article was obtained for full reading to determine conformity to the inclusion criteria. A subsample of potentially relevant articles retrieved for full-paper screening were extracted by a second author (JPS) to determine interrater agreement. If any discrepancies arose, these were resolved by discussion between authors. Interrater agreement was high (Cohen's kappa = .81).

Data Extraction and Synthesis

Data of eligible papers was extracted via standardized forms developed for this review. All available information was extracted. Identified devices which assessed where physical activity and sedentary behavior occur were tabulated to highlight the available literature in this research area and to showcase the array of measurement technologies.

Internet Search Engines

To ensure that the widest possible range of devices were included, systematic searches of Internet search engines were performed for devices and technologies that are able to measure location but may not have made their way into the published research to date. This was necessary due to the relatively slow pace of research and publication compared to the pace of technological advance (ie, new research papers may use old technology which has been surpassed by newer models). Google, Bing, and Yahoo were searched using the following key terms: RTLS (real-time locating system), GPS tracking device, RFID (radio-frequency identification) tracking, wearable camera, wearable GPS, and wearable RFID. These search terms were chosen based on the results of the academic literature searches. Specialized software was used to export the first 300 results of each search to Microsoft Excel. This ensured that the results were unaffected by the changing algorithms of search engines. Searches were completed on February 4, 2015. The retrieved website addresses were screened to determine eligibility. Only manufacturer websites were included to ensure the accuracy of the information. All other websites, including blogs and consumer review websites, were excluded. Eligible websites were then browsed for location monitoring devices. Only devices and full integrated systems which are ready to use (ie, not

bespoke) were included in an attempt to address the practicalities of deployment to assess where physical activity and sedentary time occur. The specifications of these devices were then extracted using standardized forms developed for this review. If available, specifications were obtained from device manuals. If device manuals were not available, any specifications shown on the website regarding the device were extracted. Only available information was extracted (ie, gaps in tables indicate a lack of available information). By note of caution, readers should be mindful that device characteristics, as supplied by manufacturers, are often generated under ideal conditions. Real-world pilot-testing with participants may, therefore, be required to establish real-world device characteristics.

Results

The number of research papers included and excluded at each stage of the systematic review process is shown in Figure 1. This review began with 61,009 potentially eligible papers, eventually resulting in the full inclusion of 98 papers. A further 90 papers were then identified through reference searching, citation tracking, and the searching of personal files.

A breakdown by year and technology is depicted in Figure 2. This review found 12 types of technology capable of assessing where physical activity and sedentary behavior occur. GPS was the most widely used location monitoring technology, comprising 119 (63.3%) [16,18-134] of the total 188 papers. Wearable cameras and RFID were the second- and third-most popular forms of location technology, contributing 23 (12.2%) [18,19,135-156] and 20 (10.6%) [157-177] studies, respectively, out of 188. The remaining 9 technologies each contributed a small number of studies (8 [4.3%] or less) to the total sample [178-200]. GPS has the longest history of use, initially being used within sports science in 1997. Conversely, wearable cameras and Wi-Fi-based localization technologies appear to be the most recent debut within research.

Selective details of devices used within research are shown in Table 1 (wearable cameras), Table 2 (GPS), and Table 3 (other). A complete version of Table 2 is available as Multimedia Appendix 1.

Tables 4-6 show selective characteristics of the results of the Internet search engine searches for wearable cameras, RTLS, and GPS, respectively. Complete versions of Tables 5 and 6 are available as Multimedia Appendices 2 and 3. These searches found 21 wearable cameras [201-214], 78 RTLS tags from 35 companies [215-249], and 81 GPS devices [250-286]. GPS devices were marketed for a variety of purposes, including the tracking of children by parents, elder monitoring to limit wandering, and the tracking of young drivers. RTLS companies positioned their products as suitable for asset management applications in warehouses and, to a lesser extent, equipment and patient tracking in health care settings. Wearable cameras were targeted toward extreme sports, life logging, and law enforcement applications.

Table 1. Summary of wearable camera systems used in published research to date.

Man[a]	Model	I/O[b]	BL[c], h	CR[d], mp	Dim[e], cm	Weight, kg or g	Wear site	SF[f], FPS[g]	Refs[h], notes
Natural-Point, Inc	OptiTrack-Prime 17W	Indoor	N/A[i]	1.7	12.6x12.6 x11.0	1.32 kg	N/A	30-360	[135]
Vicon	Motion capture system	Both (most in-door)	N/A	≤16	N/A	N/A	N/A	≤1000	[136,137]
Prototype	eButton	Both	~10	N/A	6.2 diameter	42 g	Pin onto shirt	10	[138-140] Not yet com-mer-cialized
Prototype	Prototype Wrist-Sense	Both	~7	N/A	N/A	N/A	Wrist	6	[141,142]
Microsoft	SenseCam (Vicon Revue)	Both	≤16	N/A	N/A	N/A	Lanyard around neck	Change in sen-sor readings	[18,19,143-155] Available as Auto-grapher
Looxcie	Looxcie 2	Both	1-4	N/A	2.31x1.70 x8.46	22 g	N/A	15/30	[156]

[a]Man: manufacturer

[b]I/O: Indoor/outdoor

[c]BL: battery life

[d]CR: camera resolution

[e]Dim: dimensions

[f]SF: sampling frequency

[g]FPS: frames per second

[h]Refs: references

[i]N/A: not applicable

Table 2. Summary of Global positioning systems used to date in published research (see Multimedia Appendix 1 for the full version of this table).

Man[a]	Model	Battery life, h, days, or weeks	Dim[b], cm or mm	Weight, g	Wear site	Cold start time, s or Hz	Storage, points, MB, or, GB	References
Garmin	Foretrex 201	15 h	8.4x4.3x 1.8 cm	78	Wrist	45 s	10,000 points	[16,20-31]
Garmin	Forerunner 305	Typically 10 h	5.3x 6.8x 1.7 cm	77	Wrist	45 s	N/A[c]	[32-35]
Garmin	Forerunner 205	10 h	53x69x 18 mm	77	Wrist	45 s	72,000 points	[25,30,42-45]
Garmin	60	N/A	N/A	N/A	Pocket of back- pack	0.5 Hz	N/A	[46,47]
Telespial Sys- tems	Trackstick II	16-36 h, 2 days- 1 week in pow- er save	11.4x3.1x 1.9 cm	N/A	N/A	Maximum of 52 s	1 MB	[52,53]
GlobalSat	DG100	20-24 h	N/A	N/A	Waist	5, 15, or 30 s	50,000 points	[25,47,54-59]
GPSports	SPI ELITE	N/A	N/A	N/A	Back har- ness	1 Hz	N/A	[60-69]
GPSports	SPI PRO	N/A	N/A	N/A	Back har- ness	5 Hz	N/A	[71-74]
GPSports	SPI 10	N/A	N/A	N/A	Back har- ness	1 Hz	N/A	[65,66,73,75-77]
Catapult Innova- tions	MinimaxX	5 h	8.8x5.0x 1.9 cm	67	Back har- ness		1 GB	[62,73,78-87]
Telespial Sys- tems	Super	4-8 days	N/A	N/A	Waist	5 or 15 s	N/A	[88,89]
Qstarz	BT1000X	42 h	72x47x 20 mm	65	Pouch on belt	35 s, 5 s, or 15 s	400,000 points	[18,19,30,59,90-104,134]
GlobalSat	BT335	25 h	N/A	N/A	Waist	30 s	N/A	[110-114]

[a]Man: manufacturer

[b]Dim: dimensions

[c]N/A: not applicable

Table 3. Summary of other measures used in published research to date.

Type of measure	Indoor/outdoor	References
Radio-frequency identification	Indoor	[157-177]
Wireless localization	Indoor	[178-183]
Technology-assisted ecological momentary assessment/experience sampling	Both	[184-191]
Integrated circuit tags	Indoor	[192,193]
Ultrasonic (Bat system)	Indoor	[194]
Cellular networks	Outdoor, but works indoor	[195]
Bluetooth	Indoor	[196]
Social media check-in	Both	[197]
Ultrasound	Indoor	[198,199]
Pedestrian dead reckoning system	Indoor	[200]

XSL•FO
RenderX

Table 4. Summary of commercially available wearable cameras unused in research to date.

Manufacturer, reference	Model	Battery life of wearable component, h or min	Dimensions, mm, in, or cm	Weight, g or oz	Wear site
Autographer [201]	N/A[a]	10 h	37.40x90.00x 22.93 mm	58 g	Clip or lanyard
Narrative (formally Memoto) [202]					
	Clip	N/A	36x36x9 mm	20 g	Clip or lanyard
	Clip 2 (released spring 2015)	N/A	N/A	N/A	N/A
MeCam [203]					
	Classic	80 min continuous	1.75x0.50 in	1 oz	Clip or necklace
	MeCam HD	60-120 min	2x2 in	2.5 oz	N/A
uCorder [204]					
	Pockito IRDC260-R	≤75 min	2.50x1.25x0.50 in	N/A	N/A
	Pockito IRDC260-B	≤75 min	2.50x1.25x0.50 in	N/A	N/A
	Pockito IRDC150	≤2 h	1.1x0.6x3.5 in	N/A	N/A
	Pockito IRDC250	≤2 h	1.1x0.6x3.5 in	N/A	N/A
ParaShoot [205]	2.1	N/A	45x4x15 mm	1.5 oz	Clip
Spy Emporium [206]	Spy hidden camera glasses	1-2 h	160x40x40 mm	N/A	Glasses/ on face
VIEVU [207]					
	VIEVU 2	2.5 h recording, 1.5 h streaming	1.90x1.90x0.75 in	2.4 oz	Clip
	LE3	≤5 h	3.00x2.10x0.85 in	2.8 oz	Clip
Panasonic [208]					
	WV-TW310L	5 h continuous	45x75x41 mm	210 g	
	WV-TW310S	5 h continuous	45x75x41 mm	160 g	
meMINI [209]	N/A	3.5 h	N/A	N/A	Lanyard
Pivothead [210]	N/A	N/A	N/A	N/A	Glasses/ on face
Nixie [211]	N/A	N/A	N/A	N/A	Wrist (detaches to become camera)
CA7CH [212]	Lightbox	N/A	38x38x10 mm	30 g	Clip
ELMO USA [213]	QBIC-MSI	2 h	2.14x2.40x1.57 in	95 g	Lanyard
Vidcie [214]	Lookout QUB	1 h (8 h with battery pack)	4.8x4.8x1.5 cm	37 g	Clip

[a]N/A: not applicable

Table 5. Summary of commercially available real-time locating systems unused in research to date (see Multimedia Appendix 2 for the full version of this table).

Manufacturer, reference	Model	Infrastructure/ method	Dimensions, mm, in, or cm	Accuracy, m, cm, or ft
Ekahau [215]				
	A4	Wi-Fi, RSSI[a] and triangulation	45x55x19 mm	1 m
	B4	Wi-Fi, RSSI and triangulation	60.0x90.0x8.5 mm	1 m
	W4	Wi-Fi, RSSI and triangulation	51.5x50.0x17.5 mm	1 m
Ubisense [216]				
	Series 7000 industrial	UWB[b], TOA[c], AOA[d]	71x64x47 mm	15 cm
	Series 7000 compact	UWB, TOA, AOA	38.0x39.0x16.5 mm	15 cm
	Series 7000 slim tag	UWB, TOA, AOA	83x42x11 mm	15 cm
	Series 700 intrinsically safe tag	UWB, TOA, AOA	38.0x39.0x25.5 mm	15 cm
	Series 9000 compact tag	UWB, TOA, AOA	38.0x39.0x16.5 mm	15 cm
Zebra [218]				
	WhereTag IV	Wi-Fi, TDOA[e]	43.7x66.0x21.3 mm	2 m
	WhereTag III	Wi-Fi, TDOA	21x66x44 mm	
Sonitor [222]	Whole system	Wi-Fi, ultrasound, RSSI	N/A[f]	1 ft
Secure Care [225]	ENVisionIT	Wi-Fi	N/A	30 cm
Mojix [226]	eLocation	Passive RFID[g]	N/A	Within 1 m
TempSys [228]	Fetch System	RF[h] and ultrasound, TDOA	N/A	0.5 m
Awarepoint [229]				
	Asset tags	ZigBee	1.8x1.3x0.5 in	Up to bay level
	Wearable tag	ZigBee	1.8x1.3x0.5 in	
Nebusens [232]	Sirius Quantum	ZigBee	22.00x32.72x5.00 mm	1 m
Essensium [233]	Mobile nodes	Wide over narrowband RF, TWR[i], TOF[j]	19.8x8.8 cm	Typically 50 cm
PLUS Location [234]				
	R1 badge tag	UWB, TDOA	38.0x78.0x9.6 mm	<1 m
	R2 tags	UWB, TDOA	87x42x10 mm	<1 m
Purelink [239]	Personnel tracking tag	RFID	85x54x4 mm	2 m
Sanitag [240]				
	Staff tag	RF, RSSI, TOF	90x61x5 mm	2.5 m
	Patient tag	RF, RSSI, TOF	43x36x10 mm	2.5 m
OpenRTLS [242]	Tag	UWB, TDOA, TWR	66x44x17 mm	10 cm

[a]RSSI: received signal strength indicator

[b]UWB: ultra wide band

[c]TOA: time of arrival

[d]AOA: angle of arrival

[e]TDOA: time difference of arrival

[f]N/A: not applicable

[g]RFID: radio-frequency identification

[h]RF: radio frequency

[i]TWR: two-way ranging

[j]TOF: time of flight

Table 6. Summary of commercially available global positioning systems unused in research to date (see Multimedia Appendix 3 for the full version of this table).

Manufacturer	Model	Battery life of wearable component, days, weeks, months, or h	Dimensions, in, cm, or mm
Trackstick [250]			
	Trackstick mini	3-14 days	3.50x1.50x0.38 in
	Trackstick II	16 h-2 days (AAA)	4.50x1.25x0.75 in
	Super Trackstick	3 days-3 weeks (AAA)	4.50x1.25x0.75 in
Trackershop-UK [251]	Pro-pod5	14-15 days	6.35x4.00x2.50 cm
Gotek7 [252]			
	Prime 1.0	10 days normal; ≤12 months with 1 update per day	N/A[a]
	Prime 2.0	15 days normal; ≤14 months (1 per day)	65x42x25 mm
Trackinapack [256]			
	Advanced	≤10 days	2.63x1.38x0.79 in
	Advanced plus	≤15 days	2.50x1.50x0.79 in
TracLogik [259]			
	Guardian GPS	100-220 hours	67.8x37.0x20.0 mm
	Guardian pro GPS	2-14 days	62.5x40.0x25.0 mm
	Covert 2000	10-15 days	61x34x31 mm
Loc8tor [261]	N/A	≤9 months in power save; 3-14 days normally	68x36x20 mm
LandAirSea [269]			
	Silvercloud realtime GPS tracker	5-6 days at 2 h per day	3.90x2.26x0.90 in
	Tracking key pro	2 weeks (4 h), 4 weeks (2 h), 6 weeks (1 h per day)	3.01x1.95x1.40 in
GTX Corp [273]			
	Prime AT	≤16 days	67x37x20 mm
	Smart sole	2-3 days	Depends on show size
Nike [276]	Sportwatch GPS	8 h with average use	1.5x10.1x0.6 in
Garmin [277]			
	Forerunner 620	6 weeks (watch) 10 h (training)	45.0x45.0x12.5 mm
	Forerunner 220	6 weeks (watch) 10 h (training)	45.0x45.0x12.5 mm
	Tactix	50 h (5 weeks in watch mode)	49x49x17 mm
	Fenix 2	20 h (5 weeks in watch mode)	49x49x17 mm
Trax [280]	Trax	1 day	38x55x10 mm
Personal GPS Trackers [282]			
	Personal GPS Tracker	≤7 days	65x40x18 mm
	Mini GPS Tracker	2-4 days	58x22x11 mm

[a]N/A: not applicable

XSL•FO

RenderX

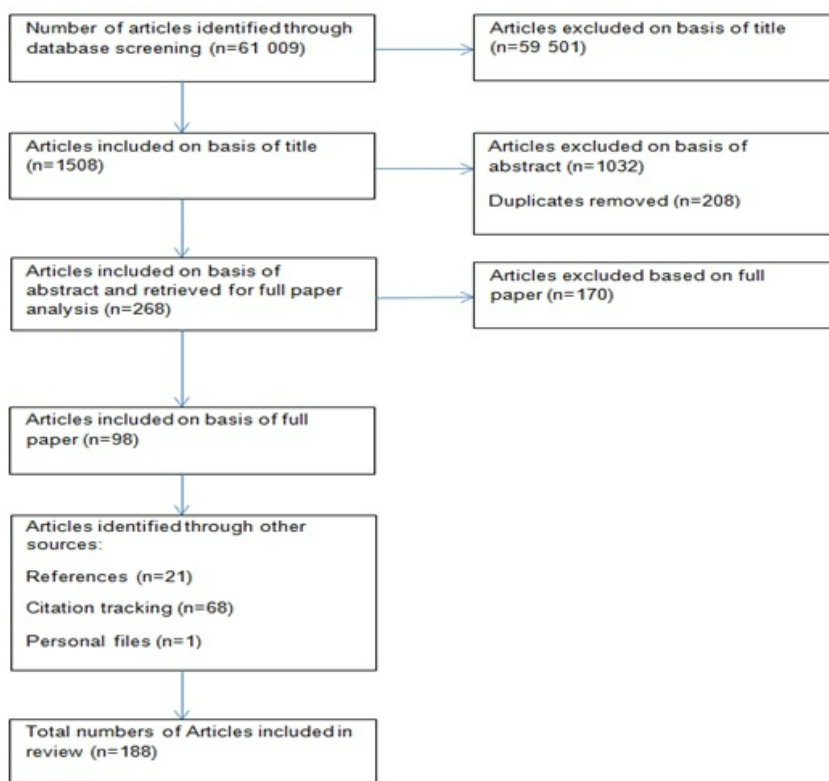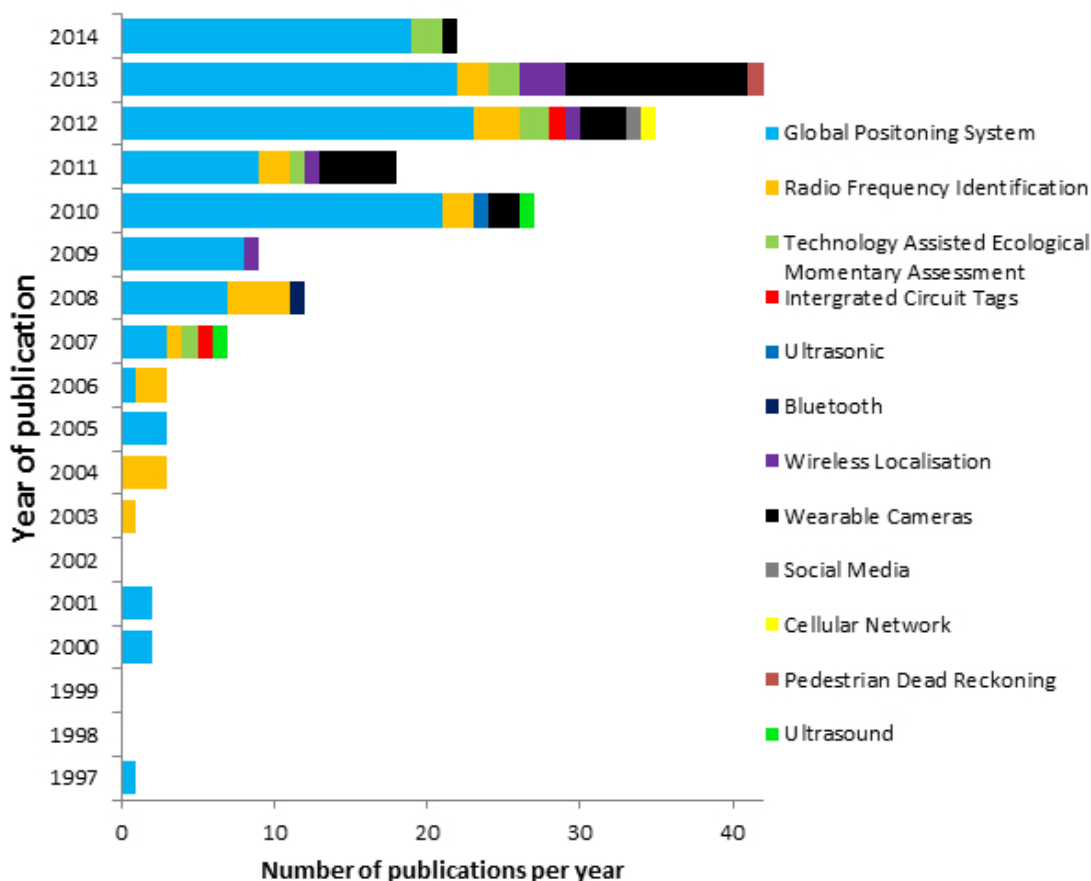**Figure 1.** Flowchart of study selection process.

Figure 2. Number of studies published each year covering different types of technology. A total of 12 kinds of technology were found during the course of this review.

Discussion

Principal Findings

This systematic review sought to identify tools which have been used, or could be modified for use, to assess where physical activity and sedentary behaviors occur. This review identified 188 research papers which used 12 different types of technology. The most widely used technology was GPS with 119 publications [16,18-134], followed by wearable cameras and RFID with 23 [18,19,135-156] and 20 [157-177] publications, respectively. The remaining 9 types of technology each contributed a small number of studies to the total sample [178-200]. However, it should be noted that a number of these were bespoke or prototype systems; this is particularly true of RFID, integrated circuit (IC) tag systems, and various communication protocols for wireless localization.

Systematic grey searches identified 21 wearable cameras [201-214], 78 RTLS tags [215-249], and 81 GPS devices [250-286]. By only including devices which are "ready to use," we sought to address the practicalities of deployment and limit the inclusion of bespoke technologies. Combined with the devices used within research papers to date, we identified a total of 263 devices. The history, principles of use, and the applications for GPS, RTLS, and wearable cameras will now be discussed in greater detail.

Global Positioning System

Originally developed by the United States Department of Defense, the GPS system consists of 24 satellites orbiting Earth. These satellites transmit signals to GPS receivers and are able to determine the location, direction, and speed of the receiver based on trilateration between three or more satellites [287]. Due to the original military application of GPS, a deliberate error was embedded into the system to reduce the risk of enemy forces using the system. This deliberate error was removed in the year 2000, thus making the system available to civilian users. The use of GPS has since proliferated into areas such as criminal offender tracking, vehicle tracking, and vehicle navigation. Such has been the widespread adoption of GPS, that the European Union is currently investing substantial amounts of money into its own satellite system to ensure it is not reliant on American satellites. Early GPS devices possessed limited battery life and memory capacity and form factors unsuitable for long periods of wear. Thus GPS devices were first used for sports applications before making their way into health research.

The earliest GPS study in a sporting domain was conducted in 1997 [132]. It was found from this initial evaluation that GPS could be used to assess human locomotion [132]. Following this early study, GPS has been used to assess movement characteristics in sports such as Australian football [66], orienteering [49], hockey [63], and rugby [72]. These studies have generally found GPS to be a suitable measure of movement parameters in sport, such as speed and distance. Physiological measures such as heart rate are often included alongside GPS to provide further data on the demands of a particular sport. These devices are often worn on the back via a custom-made vest and are, therefore, unlikely to be suitable for long-term wear. These sports studies, therefore, provide little insight into the applicability of GPS for assessing free-living physical activity.

The earliest study to use GPS to investigate free-living physical activity was conducted in 2005 [22]. The GPS units were found to provide valid and reliable measures of location when compared to a known geodetic point [22]. Following the validation of these units, a small pilot study examined the feasibility of integrating GPS, geographic information system (GIS), and accelerometer data. It was found that GPS and accelerometer data could be successfully integrated, with GPS data available for 67% of all MVPA time [22]. Accelerometer, GIS, and GPS data have since been successfully integrated in further studies to assess active commuting to school [16] and time spent outdoors after school [20].

In reviewing 24 studies which use GPS in physical activity research [288], GPS data loss was found to be highly correlated with device wear time (r=.81, *P*<.001). Common reasons for data loss include signal dropout, limited battery power, and poor protocol adherence [288]. Due to devices requiring a line of sight to the orbiting satellites, signal dropout can occur when this line of sight is broken. The necessity for GPS devices to have a line of sight to at least three orbiting satellites also results in GPS only receiving signal within certain indoor environments, such as a single-story building with a wooden roof or high-story building with large windows. Even under these circumstances, GPS is unable to determine room- or subroom-level indoor location. Participants are often required to remain stationary outside before commencing a journey to ensure that the GPS device can acquire satellite signal, failure to adhere to this can result in data loss.

While GPS can be used to successfully augment accelerometer measurement of physical activity, several shortcomings need to be addressed. There is currently no established approach to the analysis and interpretation of GPS data [287]. Guidelines and common data analysis programs for the capture and analysis of GPS data, such as the Personal Activity and Location Measurement System (PALMS), are therefore highly useful in standardizing approaches. Due to requiring a clear line of sight to orbiting satellites, GPS is most suitable for assessing outdoor location. However, up to 90% of our time is spent indoors [9,10]. The ability to assess where physical activity and sedentary time occur in an indoor environment would allow the formation of a more comprehensive behavioral profile which incorporates contextual information alongside accelerometry-measured intensity and duration.

Wireless Localization

Wireless localization technology has been commercialized under the umbrella term real-time locating systems. Used in health care [289] and warehouse environments, RTLS systems are able to assess the location of people or assets within an indoor environment. Many RTLS devices are commercially available (see Table 5 and Multimedia Appendix 2). All of these devices function on the principle of determining the location of a mobile component via the known location of fixed components, though the method of determining location and the type of fixed component vary between manufacturers. Interested readers are

referred elsewhere for detailed technological reviews of wireless localization [290-293].

The fixed components of RTLS systems also vary between RTLS manufacturers. Some manufacturers, such as AeroScout, require the installation of proprietary fixed reference points. Others, such as the Ekahau system, are able to utilize existing Wi-Fi points within buildings as fixed reference points and, therefore, do not require the installation of infrastructure. Several manufacturers also provide infrared (IR) location beacons for increased location accuracy in areas of poor signal strength. Hardware of the Ekahau RTLS system is shown in Figure 3. The location of the mobile component of the RTLS system, worn by an individual or placed on equipment, is then relayed back to software supplied with the RTLS system. This software requires a floor plan of the environment being monitored; the location of the mobile component is then viewed on this floor plan or as an x and y coordinate. RTLS systems, therefore, function in much the same manner as GPS: providing x and y coordinates rather than longitude and latitude. The manufacturers of several RTLS systems suggest that their systems are capable of handling hundreds of mobile tags simultaneously. Manufacturers state that RTLS systems are generally accurate to within 2 to 3 meters.

However, RTLS systems are not without limitations. Due to their predominant use in the tracking of patients and equipment, many RTLS systems are configured for real-time monitoring and require slight modification to generate a log of coordinates for any later integration with other data streams. At present, RTLS systems are not being used in physical activity or sedentary behavior research; therefore, the feasibility of incorporating RTLS data with accelerometry is unknown. The RTLS software requires the manual setting of the scale of the floor plan and, therefore, introduces possible human error into the system.

Despite this, RTLS could potentially be used within physical activity and sedentary behavior research to answer a number of research questions which are currently assessed via self-report methods. For example, RTLS, alongside accelerometry, could provide location information to assess whether youngsters in a daycare center are more likely to be active when they are near equipment such as a sandbox or when they are near other active youngsters. Likewise, if researchers are undertaking a standing desk intervention to reduce sitting time, participants are currently often asked to self-report how much time they spend at their desk. The amount of time the participant spends at their desk may impact any possible reduction in sitting time due to the standing desk. With RTLS, researchers would be able to objectively determine the amount of time their participants were at their standing desk and thus determine the success, or otherwise, of the intervention with greater certainty.

Determining the indoor location of physical activity and sedentary behavior via RTLS may also be an important research finding in itself. For example, within an elderly care home environment, RTLS could be used to assess whether individual residents are more sedentary alone in their bedrooms or when mixing with other residents in communal areas. Depending on the findings, some residents may then be best suited to an individual intervention focusing on bedroom-based sedentary behavior while other residents may be more suited to a group intervention focusing on communal area sedentary behavior.

Figure 3. Hardware components of the Ekahau real-time locating system (RTLS) [214]. Wearable T301W white wrist tags are shown in a charging cradle; infrared beacons are shown on either side of the charging cradle.
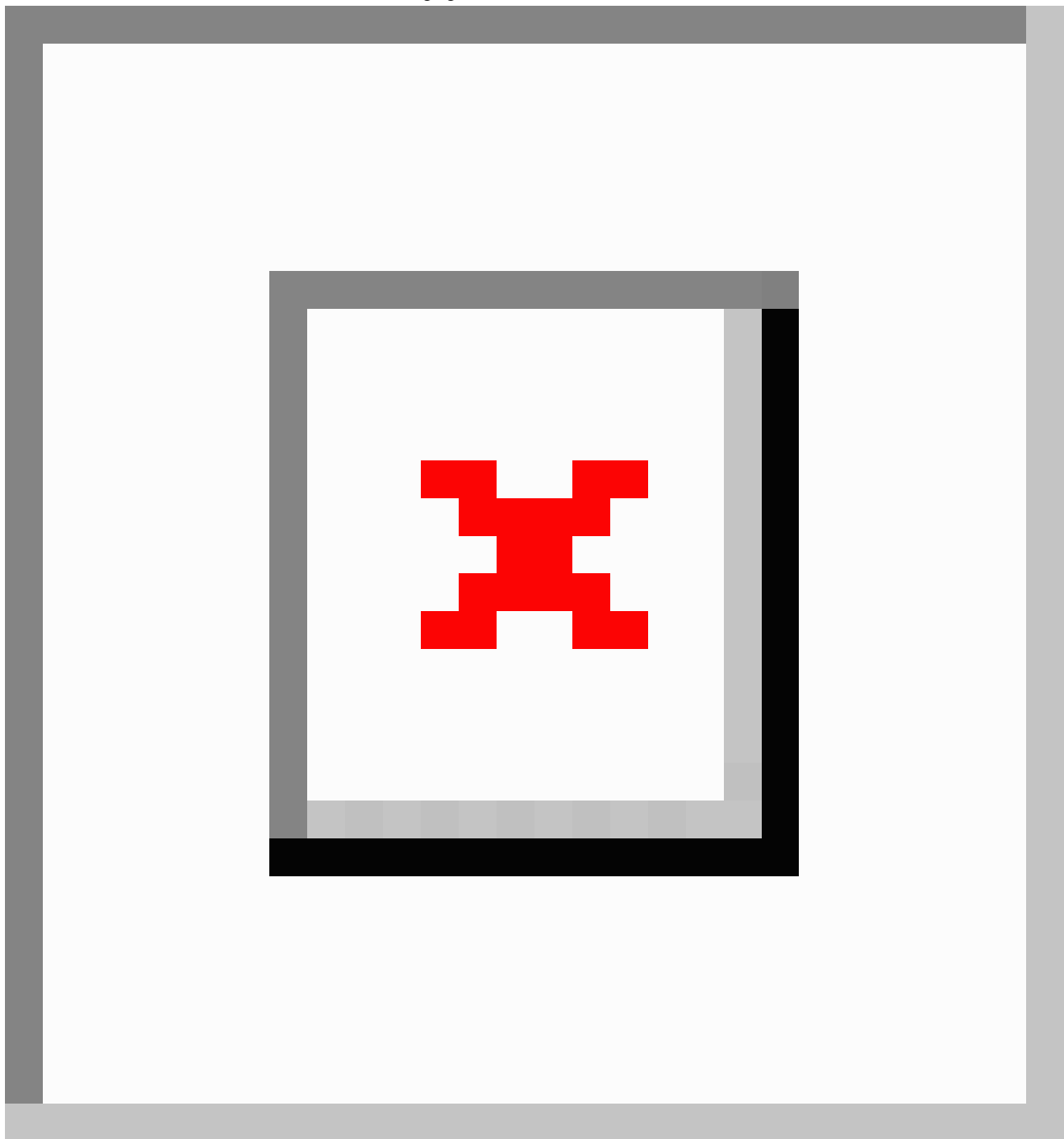
Wearable Cameras

Recent interest has accumulated in the use of wearable cameras in physical activity and sedentary behavior research, mirroring the growth of the life-logging and quantified-self communities. However, several of the wearable cameras identified in this review appear to have limited public health utility due to very short (eg, 1.5 hours) of battery life. The most popular wearable camera in a research setting is the Microsoft SenseCam. Worn on a lanyard around the neck and containing sensors such as passive infrared, accelerometer, and gyroscope, this device automatically captures a first-person picture at a frequency of approximately 20 seconds. The device has a battery life of approximately 16 hours with sufficient memory capacity to store approximately 32,000 images [294]. From initial small-scale pilot studies, it appears that images generated from wearable cameras are a feasible means of assessing active travel behavior [144,294]. Wearable cameras, therefore, provide broader contextual information; however, they can also be used to infer location. Commercially available wearable cameras, such as the Autographer, also provide GPS coordinates alongside the photograph. Two of the most popular wearable cameras are shown in Figure 4.

Unlike pure location measurement technologies such as GPS and RTLS, wearable cameras are able to provide broader contextual information based on the generated images. For example, a succession of images may show a television set.

From this, it could be identified that the participant is watching television. Likewise, a succession of images may show a group of people of a similar age to the participant which researchers may be able to classify as time spent with friends; this is important as an individual's friends may play a role in shaping physical activity behaviors [295].

Despite the encouragement offered by these initial studies, significant ethical, privacy, and analytical issues remain. There is a possibility that participants may be wearing the device during situations in which they do not wish to be photographed. To overcome this, the device allows the user to turn off the device for several minutes should they require privacy. There is also the possibility that the device may take pictures of an individual that participants encounter who does not wish to be photographed. Linked to this is the possibility that individuals may be wearing the device in situations that are unsuitable for photography, such as dropping off or picking up children from school. In an effort to overcome some of these issues [296], an ethical framework has been proposed for the use of wearable cameras in research. The framework includes the issues of informed written consent from participants, privacy and confidentiality, nonmaleficence, and the autonomy of third parties [296].

Alongside these privacy issues is the issue of data analysis. Current data analysis methods are laborious, involving the manual trawling and coding of images. For long-term monitoring this may prove to be prohibitive in the adoption of wearable cameras. Pattern recognition algorithms to semiautomate this process are available from computer scientists; however, there is a need for these to be integrated into device software in a manner which is suitable for end users. Despite these issues, wearable cameras can be used to assess where behavior occurs both indoors and outdoors and may, therefore, be able to supplement GPS to provide a greater range of contextual information.

The preceding discussion of GPS, RTLS, and wearable cameras highlights the principles, limitations, and use in physical activity and sedentary behavior research of each of these three technologies. GPS is the dominant technology used within research to date to assess where physical activity and sedentary time occur. However, the development of RTLS and wearable cameras offers the possibility to incorporate these technologies alongside GPS and accelerometry to provide a more comprehensive behavioral profile which fully elucidates the context, intensity, and duration of the behavior. The present systematic review also identified several other location monitoring technologies, such as RFID and IC tags, that are less "ready to use" than the three main technologies discussed. While these technologies, particularly RFID, may have a substantial research base behind them, there appears to be no "off the shelf" complete system which is readily purchasable for location tracking.

The ability to assess where behavior occurs in an indoor environment may be particularly elucidating for sedentary time. With the ability to assess where sedentary behavior occurs at work (eg, in a meeting room or at a desk) and at home (eg, sofa, desk, or dining table), behavioral researchers would possess a more comprehensive profile of the context in which sedentary behavior occurs, which could further illuminate the most common modes of sedentary behavior.

It is also worth briefly considering available technologies which were not included in this systematic review, largely due to a lack of wearability. Bluetooth low energy (BLE) proximity systems have recently gained in popularity in certain applications. Many of these systems are primarily aimed toward retail applications for the purpose of proximity marketing. In this scenario, small BLE beacons are placed around a retail environment. The customer, as they are perusing the store with a BLE-enabled device such as a mobile phone, then receives targeted marketing and discount offers to their phone based on their proximity to the beacons. For example, when the customer is perusing the carbonated drinks aisle in a supermarket, an offer may be sent to their phone for a particular brand of drink. These systems offer the potential to install BLE beacons within an indoor environment and determine location based on proximity to the beacons.

Of particular note is the recent miniaturization of BLE beacons to the size of a sticker, so suitably small that it may unobtrusively be attached to items such as chairs, bicycles, and sports equipment. This novel "nearables" equipment offers the potential to assess the location and type of behavior.

Figure 4. Hardware of two wearable cameras. Shown on the left is the Narrative clip [201] and on the right the Autographer [200].

Conclusions

This systematic review sought to identify tools which have been used or could be used to asses where physical activity and sedentary time occur. We identified 188 research papers, of which 119 used GPS and 23 used wearable cameras. A total of 76 location tracking devices or systems were used. Systematic Internet search engine searches found 21 wearable camera models, 78 RTLS tags, and 81 GPS devices. This gave a cumulative total of 263 location tracking devices or systems. GPS is the dominant form of location tracking used within physical activity research to date. While GPS is a valid measure of outdoor location, it is unable to be used within an indoor environment.

Recent developments in wearable cameras and RTLS systems have ensured that tools are now available which offer the potential to assess where physical activity and sedentary behaviors occur indoors. Thus, these tools can provide further contextual information, alongside GPS, when used in conjunction with measures of physical activity and sedentary behavior such as accelerometers. Issues and limitations of each technology were identified, including privacy, data analysis and interpretation, and common data processing methodologies. The integration of accelerometry, GPS, and a technology capable of assessing indoor location would provide researchers with the ability to assess the indoor and outdoor location of physical activity and sedentary behavior. Future research should, therefore, investigate the feasibility of incorporating these technologies, with particular reference to the wearability of the devices, the integration of data streams, and the generation of meaningful behavioral outcomes.

Authors' Contributions

AL, LBS, and DWE conceived and planned the review. AL completed data acquisition and drafted the manuscript. LBS, JPS, PWS, and DWE provided critical comments and insights. All authors read and approved the final manuscript.

XSL•FO

RenderX

Conflicts of Interest

None declared.

Multimedia Appendix 1

Full version of Table 2.

[PDF File (Adobe PDF File), 11KB - jmir_v17i8e192_app1.pdf]

Multimedia Appendix 2

Full version of Table 5.

[PDF File (Adobe PDF File), 11KB - jmir_v17i8e192_app2.pdf]

Multimedia Appendix 3

Full version of Table 6.

[PDF File (Adobe PDF File), 13KB - jmir_v17i8e192_app3.pdf]

References

1. Garber CE, Blissmer B, Deschenes MR, Franklin BA, Lamonte MJ, Lee I, American College of Sports Medicine. American College of Sports Medicine position stand. Quantity and quality of exercise for developing and maintaining cardiorespiratory, musculoskeletal, and neuromotor fitness in apparently healthy adults: guidance for prescribing exercise. Med Sci Sports Exerc 2011 Jul;43(7):1334-1359. [doi: 10.1249/MSS.0b013e318213fefb] [Medline: 21694556]
2. Tremblay MS, LeBlanc AG, Kho ME, Saunders TJ, Larouche R, Colley RC, et al. Systematic review of sedentary behaviour and health indicators in school-aged children and youth. Int J Behav Nutr Phys Act 2011;8:98 [FREE Full text] [doi: 10.1186/1479-5868-8-98] [Medline: 21936895]
3. Grøntved A, Hu FB. Television viewing and risk of type 2 diabetes, cardiovascular disease, and all-cause mortality: a meta-analysis. JAMA 2011 Jun 15;305(23):2448-2455 [FREE Full text] [doi: 10.1001/jama.2011.812] [Medline: 21673296]
4. Ekelund U, Luan J, Sherar LB, Esliger DW, Griew P, Cooper A, International Children's Accelerometry Database (ICAD) Collaborators. Moderate to vigorous physical activity and sedentary time and cardiometabolic risk factors in children and adolescents. JAMA 2012 Feb 15;307(7):704-712 [FREE Full text] [doi: 10.1001/jama.2012.156] [Medline: 22337681]
5. Sedentary Behaviour Research Network. Letter to the editor: standardized use of the terms "sedentary" and "sedentary behaviours". Appl Physiol Nutr Metab 2012 Jun;37(3):540-542. [doi: 10.1139/h2012-024] [Medline: 22540258]
6. Owen N, Healy GN, Matthews CE, Dunstan DW. Too much sitting: the population health science of sedentary behavior. Exerc Sport Sci Rev 2010 Jul;38(3):105-113 [FREE Full text] [doi: 10.1097/JES.0b013e3181e373a2] [Medline: 20577058]
7. Sallis J, Owen N. Physical Activity and Behavioral Medicine. Thousand Oaks, CA: Sage Publications; 1999.
8. Marshall SJ, Ramirez E. Reducing sedentary behavior: a new paradigm in physical activity promotion. Am J Lifestyle Med 2011 Feb 16;5(6):518-530. [doi: 10.1177/1559827610395487]
9. Indoor Air Division, Office of Atmospheric and Indoor Air Programs. Report to Congress on Indoor Air Quality. Volume 2. Washington, DC: United States Environmental Protection Agency; 1989 Aug 01. URL: http://nepis.epa.gov/Exe/ ZyPDF.cgi/9100LMBU.PDF?Dockey=9100LMBU.pdf [accessed 2015-07-31] [WebCite Cache ID 6aQwaHwwd]
10. European Commission, Eurostat. How Europeans Spend Their Time: Everyday Life of Women and Men. Luxembourg City, Luxembourg: Office for Official Publications of the European Communities; 2004. URL: http://bookshop.europa.eu/ en/how-europeans-spend-their-time-pbKS5804998/downloads/KS-58-04-998-EN-C/KS5804998ENC_002. pdf?FileName=KS5804998ENC_002.pdf&SKU=KS5804998ENC_PDF&CatalogueNumber=KS-58-04-998-EN-C [accessed 2015-07-31] [WebCite Cache ID 6aQww3lGr]
11. Dunstan D, Healy G, Sugiyama T, Owen N. Too much sitting and metabolic risk-has modern technology caught up with us? European Endocrinology 2010;6(1):19-23. [doi: 10.17925/EE.2010.06.00.19]
12. Preece SJ, Goulermas JY, Kenney LP, Howard D, Meijer K, Crompton R. Activity identification using body-mounted sensors--a review of classification techniques. Physiol Meas 2009 Apr;30(4):R1-R33. [doi: 10.1088/0967-3334/30/4/R01] [Medline: 19342767]
13. Intille S, Rondoni J, Kukla C, Ancona I, Bao L. A context-aware experience sampling tool. In: Proceedings of the Conference on Human Factors in Computing Systems. New York, NY: ACM; 2003 Presented at: Conference on Human Factors in Computing Systems; April 5-10, 2003; Fort Lauderdale, FL p. 972-973.
14. Hoyos CI, Jago R. Systematic review of correlates of screen-viewing among young children. Prev Med 2010 Jul;51(1):3-10. [doi: 10.1016/j.ypmed.2010.04.012] [Medline: 20417227]

15. Kozey KS, Lyden K, Hickey A, Ray EL, Fowke JH, Freedson PS, et al. Validation of a previous day recall for measuring the location and purpose of active and sedentary behaviors compared to direct observation. Int J Behav Nutr Phys Act 2014;11:12 [FREE Full text] [doi: 10.1186/1479-5868-11-12] [Medline: 24490619]

16. Cooper AR, Page AS, Wheeler BW, Griew P, Davis L, Hillsdon M, et al. Mapping the walk to school using accelerometry combined with a global positioning system. Am J Prev Med 2010 Feb;38(2):178-183. [doi: 10.1016/j.amepre.2009.10.036] [Medline: 20117574]

17. Tandon PS, Saelens BE, Zhou C, Kerr J, Christakis DA. Indoor versus outdoor time in preschoolers at child care. Am J Prev Med 2013 Jan;44(1):85-88. [doi: 10.1016/j.amepre.2012.09.052] [Medline: 23253655]

18. Lam M, Godbole S, Chen J, Oliver M, Badland H, Marshall S, et al. Measuring time spent outdoors using a wearable camera and GPS. In: Proceedings of the 4th International SenseCam and Pervasive Imaging Conference. New York, NY: ACM; 2013 Presented at: 4th International SenseCam and Pervasive Imaging Conference; November 18-19, 2013; San Diego, CA p. 1-7. [doi: 10.1145/2526667.2526668]

19. Carlson JA, Jankowska MM, Meseck K, Godbole S, Natarajan L, Raab F, et al. Validity of PALMS GPS scoring of active and passive travel compared with SenseCam. Med Sci Sports Exerc 2015 Mar;47(3):662-667. [doi: 10.1249/MSS.0000000000000446] [Medline: 25010407]

20. Cooper AR, Page AS, Wheeler BW, Hillsdon M, Griew P, Jago R. Patterns of GPS measured time outdoors after school and objective physical activity in English children: the PEACH project. Int J Behav Nutr Phys Act 2010;7:31 [FREE Full text] [doi: 10.1186/1479-5868-7-31] [Medline: 20412582]

21. Lachowycz K, Jones AP, Page AS, Wheeler BW, Cooper AR. What can global positioning systems tell us about the contribution of different types of urban greenspace to children's physical activity? Health Place 2012 May;18(3):586-594 [FREE Full text] [doi: 10.1016/j.healthplace.2012.01.006] [Medline: 22365385]

22. Rodríguez DA, Brown AL, Troped PJ. Portable global positioning units to complement accelerometry-based physical activity monitors. Med Sci Sports Exerc 2005 Nov;37(11 Suppl):S572-S581. [Medline: 16294120]

23. Rodriguez DA, Cho G, Elder JP, Conway TL, Evenson KR, Ghosh-Dastidar B, et al. Identifying walking trips from GPS and accelerometer data in adolescent females. J Phys Act Health 2012 Mar;9(3):421-431 [FREE Full text] [Medline: 21934163]

24. Wheeler BW, Cooper AR, Page AS, Jago R. Greenspace and children's physical activity: a GPS/GIS analysis of the PEACH project. Prev Med 2010 Aug;51(2):148-152. [doi: 10.1016/j.ypmed.2010.06.001] [Medline: 20542493]

25. Wieters KM, Kim J, Lee C. Assessment of wearable global positioning system units for physical activity research. J Phys Act Health 2012 Sep;9(7):913-923. [Medline: 21975729]

26. Cho G, Rodríguez DA, Evenson KR. Identifying walking trips using GPS data. Med Sci Sports Exerc 2011 Feb;43(2):365-372. [doi: 10.1249/MSS.0b013e3181ebec3c] [Medline: 20581721]

27. Southward EF, Page AS, Wheeler BW, Cooper AR. Contribution of the school journey to daily physical activity in children aged 11-12 years. Am J Prev Med 2012 Aug;43(2):201-204. [doi: 10.1016/j.amepre.2012.04.015] [Medline: 22813686]

28. Zenk SN, Schulz AJ, Odoms-Young AM, Wilbur J, Matthews S, Gamboa C, et al. Feasibility of using global positioning systems (GPS) with diverse urban adults: before and after data on perceived acceptability, barriers, and ease of use. J Phys Act Health 2012 Sep;9(7):924-934 [FREE Full text] [Medline: 21952361]

29. Zenk SN, Schulz AJ, Matthews SA, Odoms-Young A, Wilbur J, Wegrzyn L, et al. Activity space environment and dietary and physical activity behaviors: a pilot study. Health Place 2011 Sep;17(5):1150-1161 [FREE Full text] [doi: 10.1016/j.healthplace.2011.05.001] [Medline: 21696995]

30. Duncan S, Stewart TI, Oliver M, Mavoa S, MacRae D, Badland HM, et al. Portable global positioning system receivers: static validity and environmental conditions. Am J Prev Med 2013 Feb;44(2):e19-e29. [doi: 10.1016/j.amepre.2012.10.013] [Medline: 23332343]

31. Rodríguez DA, Cho G, Evenson KR, Conway TL, Cohen D, Ghosh-Dastidar B, et al. Out and about: association of the built environment with physical activity behaviors of adolescent females. Health Place 2012 Jan;18(1):55-62 [FREE Full text] [doi: 10.1016/j.healthplace.2011.08.020] [Medline: 21945085]

32. Collins P, Al-Nakeeb Y, Nevill A, Lyons M. The impact of the built environment on young people's physical activity patterns: a suburban-rural comparison using GPS. Int J Environ Res Public Health 2012 Sep;9(9):3030-3050 [FREE Full text] [doi: 10.3390/ijerph9093030] [Medline: 23202669]

33. Fjørtoft I, Löfman O, Halvorsen TK. Schoolyard physical activity in 14-year-old adolescents assessed by mobile GPS and heart rate monitoring analysed by GIS. Scand J Public Health 2010 Nov;38(5 Suppl):28-37. [doi: 10.1177/1403494810384909] [Medline: 21062837]

34. Maddison R, Jiang Y, Vander HS, Exeter D, Mhurchu CN, Dorey E. Describing patterns of physical activity in adolescents using global positioning systems and accelerometry. Pediatr Exerc Sci 2010 Aug;22(3):392-407. [Medline: 20814035]

35. Tan H, Wilson AM, Lowe J. Measurement of stride parameters using a wearable GPS and inertial measurement unit. J Biomech 2008;41(7):1398-1406. [doi: 10.1016/j.jbiomech.2008.02.021] [Medline: 18423472]

36. Duncan MJ, Mummery WK. GIS or GPS? A comparison of two methods for assessing route taken during active transport. Am J Prev Med 2007 Jul;33(1):51-53. [doi: 10.1016/j.amepre.2007.02.042] [Medline: 17572312]

XSL•FO

RenderX

37. Duncan MJ, Mummery WK, Dascombe BJ. Utility of global positioning system to measure active transport in urban areas. Med Sci Sports Exerc 2007 Oct;39(10):1851-1857. [doi: 10.1249/mss.0b013e31811ff31e] [Medline: 17909415]

38. Oreskovic NM, Blossom J, Field AE, Chiang SR, Winickoff JP, Kleinman RE. Combining global positioning system and accelerometer data to determine the locations of physical activity in children. Geospat Health 2012 May;6(2):263-272. [doi: 10.4081/gh.2012.144] [Medline: 22639128]

39. Fjørtoft I, Kristoffersen B, Sageie J. Children in schoolyards: tracking movement patterns and physical activity in schoolyards using global positioning system and heart rate monitoring. Landsc Urban Plan 2009 Dec;93(3-4):210-217. [doi: 10.1016/j.landurbplan.2009.07.008]

40. Seeger C, Welk G, Erickson M. Assessing the built environment using gps, physical activity monitors and geospatial surveys. International Journal of Geoinformatics 2009;5(1):41-48.

41. Seeger C, Welk G, Erickson S. Using global position systems (GPS) and physical activity monitors to assess the built environment. J Urban Reg Inf Syst Assoc 2008;20(2):5-12.

42. Webber SC, Porter MM. Monitoring mobility in older adults using global positioning system (GPS) watches and accelerometers: a feasibility study. J Aging Phys Act 2009 Oct;17(4):455-467. [Medline: 19940324]

43. Coombes E, van Sluijs E, Jones A. Is environmental setting associated with the intensity and duration of children's physical activity? Findings from the SPEEDY GPS study. Health Place 2013 Mar;20:62-65 [FREE Full text] [doi: 10.1016/j.healthplace.2012.11.008] [Medline: 23376730]

44. Jones AP, Coombes EG, Griffin SJ, van Sluijs E. Environmental supportiveness for physical activity in English schoolchildren: a study using Global Positioning Systems. Int J Behav Nutr Phys Act 2009;6:42 [FREE Full text] [doi: 10.1186/1479-5868-6-42] [Medline: 19615073]

45. Lee C, Li L. Demographic, physical activity, and route characteristics related to school transportation: an exploratory study. Am J Health Promot 2014;28(3 Suppl):S77-S88. [doi: 10.4278/ajhp.130430-QUAN-211] [Medline: 24380470]

46. Le Faucheur A, Abraham P, Jaquinandi V, Bouyé P, Saumet JL, Noury-Desvaux B. Study of human outdoor walking with a low-cost GPS and simple spreadsheet analysis. Med Sci Sports Exerc 2007 Sep;39(9):1570-1578. [doi: 10.1249/mss.0b013e3180cc20c7] [Medline: 17805090]

47. Noury-Desvaux B, Abraham P, Mahé G, Sauvaget T, Leftheriotis G, Le Faucheur A. The accuracy of a simple, low-cost GPS data logger/receiver to study outdoor human walking in view of health and clinical studies. PLoS One 2011;6(9):e23027 [FREE Full text] [doi: 10.1371/journal.pone.0023027] [Medline: 21931593]

48. Larsson P, Henriksson-Larsén K. Combined metabolic gas analyser and dGPS analysis of performance in cross-country skiing. J Sports Sci 2005 Aug;23(8):861-870. [doi: 10.1080/02640410400022078] [Medline: 16195038]

49. Larsson P, Henriksson-Larsén K. The use of dGPS and simultaneous metabolic measurements during orienteering. Med Sci Sports Exerc 2001 Nov;33(11):1919-1924. [Medline: 11689744]

50. Hongu N, Orr BJ, Roe DJ, Reed RG, Going SB. Global positioning system watches for estimating energy expenditure. J Strength Cond Res 2013 Nov;27(11):3216-3220. [doi: 10.1519/JSC.0b013e31828bae0f] [Medline: 23439338]

51. Nielsen RO, Cederholm P, Buist I, Sørensen H, Lind M, Rasmussen S. Can GPS be used to detect deleterious progression in training volume among runners? J Strength Cond Res 2013 Jun;27(6):1471-1478. [doi: 10.1519/JSC.0b013e3182711e3c] [Medline: 22990565]

52. Badland HM, Duncan MJ, Oliver M, Duncan JS, Mavoa S. Examining commute routes: applications of GIS and GPS technology. Environ Health Prev Med 2010 Sep;15(5):327-330 [FREE Full text] [doi: 10.1007/s12199-010-0138-1] [Medline: 21432562]

53. Oliver M, Badland H, Mavoa S, Duncan MJ, Duncan S. Combining GPS, GIS, and accelerometry: methodological issues in the assessment of location and intensity of travel behaviors. J Phys Act Health 2010 Jan;7(1):102-108. [Medline: 20231761]

54. Abraham P, Noury-Desvaux B, Gernigon M, Mahé G, Sauvaget T, Leftheriotis G, et al. The inter- and intra-unit variability of a low-cost GPS data logger/receiver to study human outdoor walking in view of health and clinical studies. PLoS One 2012;7(2):e31338 [FREE Full text] [doi: 10.1371/journal.pone.0031338] [Medline: 22363623]

55. Quigg R, Gray A, Reeder AI, Holt A, Waters DL. Using accelerometers and GPS units to identify the proportion of daily physical activity located in parks with playgrounds in New Zealand children. Prev Med 2010;50(5-6):235-240. [doi: 10.1016/j.ypmed.2010.02.002] [Medline: 20153361]

56. Kang B, Moudon AV, Hurvitz PM, Reichley L, Saelens BE. Walking objectively measured: classifying accelerometer data with GPS and travel diaries. Med Sci Sports Exerc 2013 Jul;45(7):1419-1428 [FREE Full text] [doi: 10.1249/MSS.0b013e318285f202] [Medline: 23439414]

57. Hurvitz PM, Moudon AV, Kang B, Fesinmeyer MD, Saelens BE. How far from home? The locations of physical activity in an urban U.S. setting. Prev Med 2014 Dec;69:181-186. [doi: 10.1016/j.ypmed.2014.08.034] [Medline: 25285750]

58. Brown BB, Wilson L, Tribby CP, Werner CM, Wolf J, Miller HJ, et al. Adding maps (GPS) to accelerometry data to improve study participants' recall of physical activity: a methodological advance in physical activity research. Br J Sports Med 2014 Jul;48(13):1054-1058 [FREE Full text] [doi: 10.1136/bjsports-2014-093530] [Medline: 24815545]

59. Hurvitz PM, Moudon AV, Kang B, Saelens BE, Duncan GE. Emerging technologies for assessing physical activity behaviors in space and time. Front Public Health 2014;2:2 [FREE Full text] [doi: 10.3389/fpubh.2014.00002] [Medline: 24479113]

60. Barbero-Alvarez JC, Coutts A, Granda J, Barbero-Alvarez V, Castagna C. The validity and reliability of a global positioning satellite system device to assess speed and repeated sprint ability (RSA) in athletes. J Sci Med Sport 2010 Mar;13(2):232-235. [doi: 10.1016/j.jsams.2009.02.005] [Medline: 19446495]

61. MacLeod H, Morris J, Nevill A, Sunderland C. The validity of a non-differential global positioning system for assessing player movement patterns in field hockey. J Sports Sci 2009 Jan 15;27(2):121-128. [doi: 10.1080/02640410802422181] [Medline: 19058089]

62. Duffield R, Reid M, Baker J, Spratford W. Accuracy and reliability of GPS devices for measurement of movement patterns in confined spaces for court-based sports. J Sci Med Sport 2010 Sep;13(5):523-525. [doi: 10.1016/j.jsams.2009.07.003] [Medline: 19853507]

63. Macutkiewicz D, Sunderland C. The use of GPS to evaluate activity profiles of elite women hockey players during match-play. J Sports Sci 2011 Jun;29(9):967-973. [doi: 10.1080/02640414.2011.570774] [Medline: 21604228]

64. Suárez-Arrones L, Portillo L, González-Ravé J, Muñoz V, Sanchez F. Match running performance in Spanish elite male rugby union using global positioning system. Isokinetics Exerc Sci 2012;20(2):77-83. [doi: 10.3233/IES-2012-0444]

65. Coutts AJ, Duffield R. Validity and reliability of GPS devices for measuring movement demands of team sports. J Sci Med Sport 2010 Jan;13(1):133-135. [doi: 10.1016/j.jsams.2008.09.015] [Medline: 19054711]

66. Wisbey B, Montgomery PG, Pyne DB, Rattray B. Quantifying movement demands of AFL football using GPS tracking. J Sci Med Sport 2010 Sep;13(5):531-536. [doi: 10.1016/j.jsams.2009.09.002] [Medline: 19897414]

67. Cunniffe B, Proctor W, Baker JS, Davies B. An evaluation of the physiological demands of elite rugby union using Global Positioning System tracking software. J Strength Cond Res 2009 Jul;23(4):1195-1203. [doi: 10.1519/JSC.0b013e3181a3928b] [Medline: 19528840]

68. Doğramac SN, Watsford ML, Murphy AJ. The reliability and validity of subjective notational analysis in comparison to global positioning system tracking to assess athlete movement patterns. J Strength Cond Res 2011 Mar;25(3):852-859. [doi: 10.1519/JSC.0b013e3181c69edd] [Medline: 20647953]

69. Sindall P, Lenton JP, Whytock K, Tolfrey K, Oyster ML, Cooper RA, et al. Criterion validity and accuracy of global positioning satellite and data logging devices for wheelchair tennis court movement. J Spinal Cord Med 2013 Jul;36(4):383-393 [FREE Full text] [doi: 10.1179/2045772312Y.0000000068] [Medline: 23820154]

70. Gray AJ, Jenkins D, Andrews MH, Taaffe DR, Glover ML. Validity and reliability of GPS for measuring distance travelled in field-based team sports. J Sports Sci 2010 Oct;28(12):1319-1325. [doi: 10.1080/02640414.2010.504783] [Medline: 20859825]

71. McLellan CP, Lovell DI, Gass GC. Performance analysis of elite Rugby League match play using global positioning systems. J Strength Cond Res 2011 Jun;25(6):1703-1710. [doi: 10.1519/JSC.0b013e3181ddf678] [Medline: 21358424]

72. Waldron M, Worsfold P, Twist C, Lamb K. Concurrent validity and test-retest reliability of a global positioning system (GPS) and timing gates to assess sprint performance variables. J Sports Sci 2011 Dec;29(15):1613-1619. [doi: 10.1080/02640414.2011.608703] [Medline: 22004326]

73. Petersen C, Pyne D, Portus M, Dawson B. Validity and reliability of GPS units to monitor cricket-specific movement patterns. Int J Sports Physiol Perform 2009 Sep;4(3):381-393. [Medline: 19953825]

74. McLellan CP, Lovell DI. Performance analysis of professional, semiprofessional, and junior elite rugby league match-play using global positioning systems. J Strength Cond Res 2013 Dec;27(12):3266-3274. [doi: 10.1519/JSC.0b013e31828f1d74] [Medline: 23478474]

75. Stigell E, Schantz P. Methods for determining route distances in active commuting – Their validity and reproducibility. J Transp Geogr 2011 Jul;19(4):563-574. [doi: 10.1016/j.jtrangeo.2010.06.006]

76. Edgecomb SJ, Norton KI. Comparison of global positioning and computer-based tracking systems for measuring player movement distance during Australian football. J Sci Med Sport 2006 May;9(1-2):25-32. [doi: 10.1016/j.jsams.2006.01.003] [Medline: 16580251]

77. Brewer C, Dawson B, Heasman J, Stewart G, Cormack S. Movement pattern comparisons in elite (AFL) and sub-elite (WAFL) Australian football games using GPS. J Sci Med Sport 2010 Nov;13(6):618-623. [doi: 10.1016/j.jsams.2010.01.005] [Medline: 20434398]

78. Janssen I, Sachlikidis A. Validity and reliability of intra-stroke kayak velocity and acceleration using a GPS-based accelerometer. Sports Biomech 2010 Mar;9(1):47-56. [doi: 10.1080/14763141003690229] [Medline: 20446639]

79. Portas MD, Harley JA, Barnes CA, Rush CJ. The validity and reliability of 1-Hz and 5-Hz global positioning systems for linear, multidirectional, and soccer-specific activities. Int J Sports Physiol Perform 2010 Dec;5(4):448-458. [Medline: 21266730]

80. Castellano J, Casamichana D. Heart rate and motion analysis by GPS in beach soccer. J Sports Sci Med 2010;9(1):98-103 [FREE Full text] [Medline: 24149392]

81. Jennings D, Cormack S, Coutts AJ, Boyd L, Aughey RJ. The validity and reliability of GPS units for measuring distance in team sport specific running patterns. Int J Sports Physiol Perform 2010 Sep;5(3):328-341. [Medline: 20861523]

82. Johnston RJ, Watsford ML, Pine MJ, Spurrs RW, Murphy AJ, Pruyn EC. The validity and reliability of 5-Hz global positioning system units to measure team sport movement demands. J Strength Cond Res 2012 Mar;26(3):758-765. [doi: 10.1519/JSC.0b013e318225f161] [Medline: 22310508]

XSL•FO
RenderX

83. Varley MC, Fairweather IH, Aughey RJ. Validity and reliability of GPS for measuring instantaneous velocity during acceleration, deceleration, and constant motion. J Sports Sci 2012;30(2):121-127. [doi: 10.1080/02640414.2011.627941] [Medline: 22122431]

84. Molinos Domene A. Evaluation of movement and physiological demands of full-back and center-back soccer players using global positioning systems. Journal of Human Sport and Exercise 2013;8(4):1015-1028. [doi: 10.4100/jhse.2013.84.12]

85. White AD, MacFarlane N. Time-on-pitch or full-game GPS analysis procedures for elite field hockey? Int J Sports Physiol Perform 2013 Sep;8(5):549-555. [Medline: 23412758]

86. Dwyer DB, Gabbett TJ. Global positioning system data analysis: velocity ranges and a new definition of sprinting for field sport athletes. J Strength Cond Res 2012 Mar;26(3):818-824. [doi: 10.1519/JSC.0b013e3182276555] [Medline: 22310509]

87. O'Hara JP, Brightmore A, Till K, Mitchell I, Cummings S, Cooke CB. Evaluation of movement and physiological demands of rugby league referees using global positioning systems tracking. Int J Sports Med 2013 Sep;34(9):825-831. [doi: 10.1055/s-0033-1333694] [Medline: 23444093]

88. McMinn D, Rowe DA, Cuk I. Evaluation of thetrackstick super GPS tracker for use in walking research. Res Q Exerc Sport 2012 Mar;83(1):108-113. [doi: 10.1080/02701367.2012.10599831] [Medline: 22428418]

89. McMinn D, Oreskovic NM, Aitkenhead MJ, Johnston DW, Murtagh S, Rowe DA. The physical environment and health-enhancing activity during the school commute: global positioning system, geographical information systems and accelerometry. Geospat Health 2014 May;8(2):569-572. [doi: 10.4081/gh.2014.46] [Medline: 24893034]

90. Kerr J, Marshall S, Godbole S, Neukam S, Crist K, Wasilenko K, et al. The relationship between outdoor activity and health in older adults using GPS. Int J Environ Res Public Health 2012 Dec;9(12):4615-4625 [FREE Full text] [Medline: 23330225]

91. Christian WJ. Using geospatial technologies to explore activity-based retail food environments. Spat Spatiotemporal Epidemiol 2012 Dec;3(4):287-295. [doi: 10.1016/j.sste.2012.09.001] [Medline: 23149325]

92. Dessing D, Pierik FH, Sterkenburg RP, van Dommelen P, Maas J, de Vries SI. Schoolyard physical activity of 6-11 year old children assessed by GPS and accelerometry. Int J Behav Nutr Phys Act 2013;10:97 [FREE Full text] [doi: 10.1186/1479-5868-10-97] [Medline: 23945145]

93. Epstein DH, Tyburski M, Craig IM, Phillips KA, Jobes ML, Vahabzadeh M, et al. Real-time tracking of neighborhood surroundings and mood in urban drug misusers: application of a new method to study behavior in its geographical context. Drug Alcohol Depend 2014 Jan 1;134:22-29 [FREE Full text] [doi: 10.1016/j.drugalcdep.2013.09.007] [Medline: 24332365]

94. Klinker CD, Schipperijn J, Christian H, Kerr J, Ersbøll AK, Troelsen J. Using accelerometers and global positioning system devices to assess gender and age differences in children's school, transport, leisure and home based physical activity. Int J Behav Nutr Phys Act 2014;11:8 [FREE Full text] [doi: 10.1186/1479-5868-11-8] [Medline: 24457029]

95. O'Connor TM, Cerin E, Robles J, Lee RE, Kerr J, Butte N, et al. Feasibility study to objectively assess activity and location of Hispanic preschoolers: a short communication. Geospat Health 2013 May;7(2):375-380. [doi: 10.4081/gh.2013.94] [Medline: 23733298]

96. Evenson KR, Wen F, Hillier A, Cohen DA. Assessing the contribution of parks to physical activity using global positioning system and accelerometry. Med Sci Sports Exerc 2013 Oct;45(10):1981-1987 [FREE Full text] [doi: 10.1249/MSS.0b013e318293330e] [Medline: 23531716]

97. Robinson AI, Oreskovic NM. Comparing self-identified and census-defined neighborhoods among adolescents using GPS and accelerometer. Int J Health Geogr 2013;12:57 [FREE Full text] [doi: 10.1186/1476-072X-12-57] [Medline: 24325342]

98. Prins RG, Pierik F, Etman A, Sterkenburg RP, Kamphuis CB, van Lenthe FJ. How many walking and cycling trips made by elderly are beyond commonly used buffer sizes: results from a GPS study. Health Place 2014 May;27:127-133. [doi: 10.1016/j.healthplace.2014.01.012] [Medline: 24603010]

99. Dessing D, de Vries SI, Graham JM, Pierik FH. Active transport between home and school assessed with GPS: a cross-sectional study among Dutch elementary school children. BMC Public Health 2014;14:227 [FREE Full text] [doi: 10.1186/1471-2458-14-227] [Medline: 24597513]

100. Harrison F, Burgoine T, Corder K, van Sluijs E, Jones A. How well do modelled routes to school record the environments children are exposed to? A cross-sectional comparison of GIS-modelled and GPS-measured routes to school. Int J Health Geogr 2014;13:5 [FREE Full text] [doi: 10.1186/1476-072X-13-5] [Medline: 24529075]

101. Klinker CD, Schipperijn J, Kerr J, Ersbøll AK, Troelsen J. Context-specific outdoor time and physical activity among school-children across gender and age: using accelerometers and GPS to advance methods. Front Public Health 2014;2:20 [FREE Full text] [doi: 10.3389/fpubh.2014.00020] [Medline: 24653983]

102. Ellis K, Godbole S, Marshall S, Lanckriet G, Staudenmayer J, Kerr J. Identifying active travel behaviors in challenging environments using GPS, accelerometers, and machine learning algorithms. Front Public Health 2014;2:36 [FREE Full text] [doi: 10.3389/fpubh.2014.00036] [Medline: 24795875]

103. Moore HJ, Nixon CA, Lake AA, Douthwaite W, O'Malley CL, Pedley CL, et al. The environment can explain differences in adolescents' daily physical activity levels living in a deprived urban area: cross-sectional study using accelerometry, GPS, and focus groups. J Phys Act Health 2014 Nov;11(8):1517-1524. [doi: 10.1123/jpah.2012-0420] [Medline: 24733145]

104. Panter J, Costa S, Dalton A, Jones A, Ogilvie D. Development of methods to objectively identify time spent using active and motorised modes of travel to work: how do self-reported measures compare? Int J Behav Nutr Phys Act 2014;11:116 [FREE Full text] [doi: 10.1186/s12966-014-0116-x] [Medline: 25231500]

XSL•FO
RenderX

105. Terrier P, Ladetto Q, Merminod B, Schutz Y. High-precision satellite positioning system as a new tool to study the biomechanics of human locomotion. J Biomech 2000 Dec;33(12):1717-1722. [Medline: 11006399]

106. Terrier P, Ladetto Q, Merminod B, Schutz Y. Measurement of the mechanical power of walking by satellite positioning system (GPS). Med Sci Sports Exerc 2001 Nov;33(11):1912-1918. [Medline: 11689743]

107. Troped PJ, Oliveira MS, Matthews CE, Cromley EK, Melly SJ, Craig BA. Prediction of activity mode with global positioning system and accelerometer data. Med Sci Sports Exerc 2008 May;40(5):972-978. [doi: 10.1249/MSS.0b013e318164c407] [Medline: 18408598]

108. Troped PJ, Wilson JS, Matthews CE, Cromley EK, Melly SJ. The built environment and location-based physical activity. Am J Prev Med 2010 Apr;38(4):429-438 [FREE Full text] [doi: 10.1016/j.amepre.2009.12.032] [Medline: 20307812]

109. Krenn PJ, Oja P, Titze S. Route choices of transport bicyclists: a comparison of actually used and shortest routes. Int J Behav Nutr Phys Act 2014;11(1):31 [FREE Full text] [doi: 10.1186/1479-5868-11-31] [Medline: 24597725]

110. Almanza E, Jerrett M, Dunton G, Seto E, Pentz MA. A study of community design, greenness, and physical activity in children using satellite, GPS and accelerometer data. Health Place 2012 Jan;18(1):46-54 [FREE Full text] [doi: 10.1016/j.healthplace.2011.09.003] [Medline: 22243906]

111. Dunton GF, Liao Y, Almanza E, Jerrett M, Spruijt-Metz D, Chou C, et al. Joint physical activity and sedentary behavior in parent-child pairs. Med Sci Sports Exerc 2012 Aug;44(8):1473-1480 [FREE Full text] [doi: 10.1249/MSS.0b013e31825148e9] [Medline: 22367744]

112. Dunton GF, Liao Y, Almanza E, Jerrett M, Spruijt-Metz D, Pentz MA. Locations of joint physical activity in parent-child pairs based on accelerometer and GPS monitoring. Ann Behav Med 2013 Feb;45 Suppl 1:S162-S172 [FREE Full text] [doi: 10.1007/s12160-012-9417-y] [Medline: 23011914]

113. Jerrett M, Almanza E, Davies M, Wolch J, Dunton G, Spruitj-Metz D, et al. Smart growth community design and physical activity in children. Am J Prev Med 2013 Oct;45(4):386-392. [doi: 10.1016/j.amepre.2013.05.010] [Medline: 24050413]

114. Dunton GF, Almanza E, Jerrett M, Wolch J, Pentz MA. Neighborhood park use by children: use of accelerometry and global positioning systems. Am J Prev Med 2014 Feb;46(2):136-142 [FREE Full text] [doi: 10.1016/j.amepre.2013.10.009] [Medline: 24439346]

115. Paz-Soldan VA, Reiner RC, Morrison AC, Stoddard ST, Kitron U, Scott TW, et al. Strengths and weaknesses of Global Positioning System (GPS) data-loggers and semi-structured interviews for capturing fine-scale human mobility: findings from Iquitos, Peru. PLoS Negl Trop Dis 2014 Jun;8(6):e2888 [FREE Full text] [doi: 10.1371/journal.pntd.0002888] [Medline: 24922530]

116. Feng T, Timmermans HJ. Transportation mode recognition using GPS and accelerometer data. Transp Res Part C Emerg Technol 2013 Dec;37:118-130. [doi: 10.1016/j.trc.2013.09.014]

117. Neven A, Janssens D, Alders G, Wets G, Van Wijmeersch B, Feys P. Documenting outdoor activity and travel behaviour in persons with neurological conditions using travel diaries and GPS tracking technology: a pilot study in multiple sclerosis. Disabil Rehabil 2013 Sep;35(20):1718-1725. [doi: 10.3109/09638288.2012.751137] [Medline: 23343357]

118. Clark RA, Weragoda N, Paterson K, Telianidis S, Williams G. A pilot investigation using global positioning systems into the outdoor activity of people with severe traumatic brain injury. J Neuroeng Rehabil 2014;11:37 [FREE Full text] [doi: 10.1186/1743-0003-11-37] [Medline: 24645752]

119. Doherty ST, Oh P. A multi-sensor monitoring system of human physiology and daily activities. Telemed J E Health 2012 Apr;18(3):185-192. [doi: 10.1089/tmj.2011.0138] [Medline: 22480300]

120. MacLellan G, Baillie L. Development of a location and movement monitoring system to quantify physical activity. In: Proceedings of the Conference on Human Factors in Computing Systems. New York, NY: ACM; 2008 Presented at: Conference on Human Factors in Computing Systems; April 5-10, 2008; Florence, Italy p. 2889-2894. [doi: 10.1145/1358628.1358779]

121. Minarik T, Hofer C, Schrempf A. A novel activity monitoring device for home rehabilitation applications. In: Proceedings of the 9th IASTED International Conference on Biomedical Engineering. 2012 Presented at: 9th IASTED International Conference on Biomedical Engineering; February 15-17, 2012; Innsbruck, Austria p. 400-407. [doi: 10.2316/P.2012.764-120]

122. Li I, Dey AK, Forlizzi J. Using context to reveal factors that affect physical activity. ACM Trans Comput Hum Interact 2012 Mar 01;19(1):1-21. [doi: 10.1145/2147783.2147790]

123. Rainham DG, Bates CJ, Blanchard CM, Dummer TJ, Kirk SF, Shearer CL. Spatial classification of youth physical activity patterns. Am J Prev Med 2012 May;42(5):e87-e96. [doi: 10.1016/j.amepre.2012.02.011] [Medline: 22516507]

124. Schenk AK, Witbrodt BC, Hoarty CA, Carlson RH, Goulding EH, Potter JF, et al. Cellular telephones measure activity and lifespace in community-dwelling adults: proof of principle. J Am Geriatr Soc 2011 Feb;59(2):345-352 [FREE Full text] [doi: 10.1111/j.1532-5415.2010.03267.x] [Medline: 21288235]

125. Schutz Y, Herren R. Assessment of speed of human locomotion using a differential satellite global positioning system. Med Sci Sports Exerc 2000 Mar;32(3):642-646. [Medline: 10731007]

126. Sposaro F, Danielson J, Tyson G. iWander: An Android application for dementia patients. Conf Proc IEEE Eng Med Biol Soc 2010;2010:3875-3878. [doi: 10.1109/IEMBS.2010.5627669] [Medline: 21097072]

127. Wiehe SE, Carroll AE, Liu GC, Haberkorn KL, Hoch SC, Wilson JS, et al. Using GPS-enabled cell phones to track the travel patterns of adolescents. Int J Health Geogr 2008;7:22 [FREE Full text] [doi: 10.1186/1476-072X-7-22] [Medline: 18495025]

128. Duncan JS, Badland HM, Schofield G. Combining GPS with heart rate monitoring to measure physical activity in children: a feasibility study. J Sci Med Sport 2009 Sep;12(5):583-585. [doi: 10.1016/j.jsams.2008.09.010] [Medline: 19036637]

129. Rainham D, Krewski D, McDowell I, Sawada M, Liekens B. Development of a wearable global positioning system for place and health research. Int J Health Geogr 2008;7:59 [FREE Full text] [doi: 10.1186/1476-072X-7-59] [Medline: 19032783]

130. Townshend AD, Worringham CJ, Stewart IB. Assessment of speed and position during human locomotion using nondifferential GPS. Med Sci Sports Exerc 2008 Jan;40(1):124-132. [doi: 10.1249/mss.0b013e3181590bc2] [Medline: 18091013]

131. Witte TH, Wilson AM. Accuracy of WAAS-enabled GPS for the determination of position and speed over ground. J Biomech 2005 Aug;38(8):1717-1722. [doi: 10.1016/j.jbiomech.2004.07.028] [Medline: 15958230]

132. Schutz Y, Chambaz A. Could a satellite-based navigation system (GPS) be used to assess the physical activity of individuals on earth? Eur J Clin Nutr 1997 May;51(5):338-339. [Medline: 9152686]

133. Moayeri PG, Troped P, Evans J. Environment feature extraction and classification for Context aware Physical Activity monitoring. In: Proceedings of the IEEE Sensors Applications Symposium (SAS). 2013 Presented at: IEEE Sensors Applications Symposium (SAS); February 19-21, 2013; Galveston, TX p. 123-128. [doi: 10.1109/SAS.2013.6493570]

134. Oreskovic NM, Perrin JM, Robinson AI, Locascio JJ, Blossom J, Chen ML, et al. Adolescents' use of the built environment for physical activity. BMC Public Health 2015;15:251 [FREE Full text] [doi: 10.1186/s12889-015-1596-6] [Medline: 25880654]

135. Zhu C, Sheng W. Motion- and location-based online human daily activity recognition. Pervasive Mob Comput 2011 Apr;7(2):256-269. [doi: 10.1016/j.pmcj.2010.11.004]

136. Zhu C, Sheng W. Real-time recognition of complex daily activities using dynamic Bayesian network. In: Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems. Washington, DC: IEEE; 2011 Presented at: IEEE/RSJ International Conference on Intelligent Robots and Systems; September 25-30, 2011; San Francisco, CA p. 3395-3400.

137. Zhu C, Sheng W. Realtime recognition of complex human daily activities using human motion and location data. IEEE Trans Biomed Eng 2012 Sep;59(9):2422-2430. [doi: 10.1109/TBME.2012.2190602] [Medline: 22434793]

138. Zhang H, Li L, Jia W, Fernstrom JD, Sclabassi RJ, Sun M. Recognizing physical activity from ego-motion of a camera. Conf Proc IEEE Eng Med Biol Soc 2010;2010:5569-5572. [doi: 10.1109/IEMBS.2010.5626794] [Medline: 21096480]

139. Li L, Zhang H, Jia W, Mao Z, You Y, Sun M. Indirect activity recognition using a target-mounted camera. In: Proceedings of the 4th International Congress on Image and Signal Processing (CISP). Washington, DC: IEEE; 2011 Presented at: 4th International Congress on Image and Signal Processing (CISP); October 15-17, 2011; Shanghai, China p. 487-491.

140. Zhang H, Li L, Jia W, Fernstrom JD, Sclabassi RJ, Mao Z, et al. Physical activity recognition based on motion in images acquired by a wearable camera. Neurocomputing 2011 Jun 1;74(12-13):2184-2192 [FREE Full text] [doi: 10.1016/j.neucom.2011.02.014] [Medline: 21779142]

141. Maekawa T, Yanagisawa Y, Kishino Y, Ishiguro K, Kamei K, Sakurai Y, et al. Object-based activity recognition with heterogeneous sensors on wrist. In: Proceedings of the 8th International Conference, Pervasive Computing. Berlin, Germany: Springer Berlin Heidelberg; 2010 Presented at: 8th International Conference, Pervasive Computing; May 17-20, 2010; Helsinki, Finland p. 246-264. [doi: 10.1007/978-3-642-12654-3_15]

142. Maekawa T, Kishino Y, Yanagisawa Y, Sakurai Y. WristSense: wrist-worn sensor device with camera for daily activity recognition. In: Proceedings of the IEEE International Conference on Pervasive Computing and Communications Workshops. 2012 Presented at: IEEE International Conference on Pervasive Computing and Communications Workshops; March 19-23, 2012; Lugano, Switzerland p. 510-512. [doi: 10.1109/PerComW.2012.6197551]

143. Mavoa S, Oliver M, Kerr J, Doherty A, Witten K. Using SenseCam images to assess the environment. In: Proceedings of the 4th International SenseCam and Pervasive Imaging Conference. New York, NY: ACM; 2013 Presented at: 4th International SenseCam and Pervasive Imaging Conference; November 18-19, 2013; San Diego, CA p. 84-85. [doi: 10.1145/2526667.2526683]

144. Kelly P, Doherty AR, Hamilton A, Matthews A, Batterham AM, Nelson M, et al. Evaluating the feasibility of measuring travel to school using a wearable camera. Am J Prev Med 2012 Nov;43(5):546-550 [FREE Full text] [doi: 10.1016/j.amepre.2012.07.027] [Medline: 23079179]

145. Wang P, Smeaton AF. Using visual lifelogs to automatically characterize everyday activities. Inf Sci 2013 May;230:147-161. [doi: 10.1016/j.ins.2012.12.028]

146. Kelly P, Doherty A, Berry E, Hodges S, Batterham AM, Foster C. Can we use digital life-log images to investigate active and sedentary travel behaviour? Results from a pilot study. Int J Behav Nutr Phys Act 2011;8:44 [FREE Full text] [doi: 10.1186/1479-5868-8-44] [Medline: 21599935]

147. Doherty AR, Kelly P, Kerr J, Marshall S, Oliver M, Badland H, et al. Using wearable cameras to categorise type and context of accelerometer-identified episodes of physical activity. Int J Behav Nutr Phys Act 2013;10:22 [FREE Full text] [doi: 10.1186/1479-5868-10-22] [Medline: 23406270]

148. Kerr J, Marshall SJ, Godbole S, Chen J, Legge A, Doherty AR, et al. Using the SenseCam to improve classifications of sedentary behavior in free-living settings. Am J Prev Med 2013 Mar;44(3):290-296. [doi: 10.1016/j.amepre.2012.11.004] [Medline: 23415127]

149. Marinac C, Merchant G, Godbole S, Chen J, Kerr J, Clark B, et al. The feasibility of using SenseCams to measure the type and context of daily sedentary behaviors. In: Proceedings of the 4th International SenseCam and Pervasive Imaging Conference. New York, NY: ACM; 2013 Presented at: 4th International SenseCam and Pervasive Imaging Conference; November 18-19, 2013; San Diego, CA p. 42-49. [doi: 10.1145/2526667.2526674]

150. Oliver M, Doherty AR, Kelly P, Badland HM, Mavoa S, Shepherd J, et al. Utility of passive photography to objectively audit built environment features of active transport journeys: an observational study. Int J Health Geogr 2013;12:20 [FREE Full text] [doi: 10.1186/1476-072X-12-20] [Medline: 23575288]

151. Sheats J, Winter S, Padilla-Romero P, Goldman-Rosas L, Grieco L, King A. Comparison of passive versus active photo capture of built environment features by technology naive latinos using the sensecam and stanford healthy neighbourhood discovery tool. In: Proceedings of the 4th International SenseCam and Pervasive Imaging Conference. New York, NY: ACM; 2013 Presented at: 4th International SenseCam and Pervasive Imaging Conference; November 18-19, 2013; San Diego, CA p. 8-15. [doi: 10.1145/2526667.2526669]

152. O'Loughlin G, Cullen SJ, McGoldrick A, O'Connor S, Blain R, O'Malley S, et al. Using a wearable camera to increase the accuracy of dietary analysis. Am J Prev Med 2013 Mar;44(3):297-301. [doi: 10.1016/j.amepre.2012.11.007] [Medline: 23415128]

153. Barr M, Signal L, Jenkin G, Smith M. Using SenseCam to capture children's exposure to food marketing: a feasibility study. In: Proceedings of the 4th International SenseCam and Pervasive Imaging Conference. New York, NY: ACM; 2013 Presented at: 4th International SenseCam and Pervasive Imaging Conference; November 18-19, 2013; San Diego, CA p. 50-51. [doi: 10.1145/2526667.2526675]

154. Thomaz E, Parnami A, Essa I, Abowd G. Feasibility of identifying eating moments from first-person images leveraging human computation. In: Proceedings of the 4th International SenseCam and Pervasive Imaging Conference. New York, NY: ACM; 2013 Presented at: 4th International SenseCam and Pervasive Imaging Conference; November 18-19, 2013; San Diego, CA p. 26-33. [doi: 10.1145/2526667.2526672]

155. Gemming L, Doherty A, Kelly P, Utter J, Ni Mhurchu C. Feasibility of a SenseCam-assisted 24-h recall to reduce under-reporting of energy intake. Eur J Clin Nutr 2013 Oct;67(10):1095-1099. [doi: 10.1038/ejcn.2013.156] [Medline: 24002044]

156. Zariffa J, Popovic MR. Hand contour detection in wearable camera video using an adaptive histogram region of interest. J Neuroeng Rehabil 2013;10:114 [FREE Full text] [doi: 10.1186/1743-0003-10-114] [Medline: 24354542]

157. Hong Y, Kim I, Ahn S, Kim H. Activity recognition using wearable sensors for elder care. In: Proceedings of the 2nd International Conference on Future Generation Communication and Networking (FGCN). Washington, DC: IEEE; 2008 Presented at: 2nd International Conference on Future Generation Communication and Networking (FGCN); December 13-15, 2008; Hainan Island, China p. 302-305. [doi: 10.1109/FGCN.2008.165]

158. House S, Connell S, Milligan I, Austin D, Hayes TL, Chiang P. Indoor localization using pedestrian dead reckoning updated with RFID-based fiducials. In: Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). Washington, DC: IEEE; 2011 Presented at: Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC); August 30-September 3, 2011; Boston, MA p. 7598-7601. [doi: 10.1109/IEMBS.2011.6091873]

159. Hsu C, Chen J. A novel sensor-assisted RFID-based indoor tracking system for the elderly living alone. Sensors (Basel) 2011;11(11):10094-10113 [FREE Full text] [doi: 10.3390/s111110094] [Medline: 22346631]

160. Lin C, Hsu Y. IPARS: Intelligent Portable Activity Recognition System via Everyday Objects, Human Movements, and Activity Duration. Palo Alto, CA: American Association for Artificial Intelligence; 2006. URL: https://www.aaai.org/Papers/Workshops/2006/WS-06-13/WS06-13-007.pdf [accessed 2015-07-31] [WebCite Cache ID 6aR4aDAzo]

161. Lin Y, Su M, Chen S, Wang S, Lin C, Chen H. A study of ubiquitous monitor with RFID in an elderly nursing home. In: Proceedings of the International Conference on Multimedia and Ubiquitous Engineering (MUE). Washington, DC: IEEE; 2007 Presented at: International Conference on Multimedia and Ubiquitous Engineering (MUE); April 26-28, 2007; Seoul, South Korea p. 336-340. [doi: 10.1109/MUE.2007.55]

162. Stikic M, Huỳnh T, Laerhoven K, Schiele B. ADL recognition based on the combination of RFID and accelerometer sensing. In: Proceedings of the 2nd International Conference on Pervasive Computing Technologies for Healthcare. Washington, DC: IEEE; 2008 Presented at: 2nd International Conference on Pervasive Computing Technologies for Healthcare; January 30-February 1, 2008; Tampere, Finland p. 258-263. [doi: 10.1109/PCTHEALTH.2008.4571084]

163. Morton T, Weeks A, House S, Chiang P, Scaffidi C. Location and activity tracking with the cloud. In: Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). Washington, CD:

IEEE; 2012 Presented at: Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC); August 28-September 1, 2012; San Diego, CA p. 5846-5849. [doi: 10.1109/EMBC.2012.6347323]

164. Jin G, Lu X, Park M. An indoor localization mechanism using active RFID tag. In: Proceedings of the IEEE International Conference on Sensor Networks, Ubiquitous, and Trustworthy Computing. 2006 Presented at: IEEE International Conference on Sensor Networks, Ubiquitous, and Trustworthy Computing; June 5-7, 2006; Taichung, Taiwan p. 4. [doi: 10.1109/SUTC.2006.1636157]

165. Ni L, Liu Y, Lau Y, Patil A. LANDMARC: indoor location sensing using active RFID. In: Proceedings of the 1st IEEE International Conference on Pervasive Computing and Communications. Washington, DC: IEEE; 2004 Presented at: 1st IEEE International Conference on Pervasive Computing and Communications; March 26, 2003; Fort Worth, TX p. 701-710. [doi: 10.1109/PERCOM.2003.1192765]

166. Shiraishi T, Komuro N, Ueda H, Kasai H, Tsuboi T. Indoor location estimation technique using UHF band RFID. In: Proceedings of the International Conference on Information Networking. Washington, DC: IEEE; 2008 Presented at: International Conference on Information Networking; January 23-25, 2008; Busan, South Korea p. 1-5. [doi: 10.1109/ICOIN.2008.4472808]

167. Tesoriero R, Tebar R, Gallud J, Lozano M, Penichet V. Improving location awareness in indoor spaces using RFID technology. Expert Syst Appl 2010 Jan;37(1):894-898. [doi: 10.1016/j.eswa.2009.05.062]

168. Bouchard K, Fortin-Simard D, Gaboury S, Bouchard B, Bouzouane A. Accurate RFID trilateration to learn and recognize spatial activities in smart environment. International Journal of Distributed Sensor Networks 2013;2013:1-16. [doi: 10.1155/2013/936816]

169. Smith L, Ucci M, Marmot A, Spinney R, Laskowski M, Sawyer A, et al. Active buildings: modelling physical activity and movement in office buildings. An observational study protocol. BMJ Open 2013;3(11):e004103 [FREE Full text] [doi: 10.1136/bmjopen-2013-004103] [Medline: 24227873]

170. Ni L, Liu Y, Lau Y, Patil A. LANDMARC: indoor location sensing using active RFID. Wireless Networks 2004;10(6):701-710. [doi: 10.1109/PERCOM.2003.1192765]

171. Kim S, Jeong Y, Park S. RFID-based indoor location tracking to ensure the safety of the elderly in smart home environments. Pers Ubiquit Comput 2013;17(8):1699-1707. [doi: 10.1007/s00779-012-0604-4]

172. Almudevar A, Leibovici A, Tentler A. Home monitoring using wearable radio frequency transmitters. Artif Intell Med 2008 Feb;42(2):109-120. [doi: 10.1016/j.artmed.2007.11.002] [Medline: 18215512]

173. Frencken WG, Lemmink KA, Delleman NJ. Soccer-specific accuracy and validity of the local position measurement (LPM) system. J Sci Med Sport 2010 Nov;13(6):641-645. [doi: 10.1016/j.jsams.2010.04.003] [Medline: 20594910]

174. Ogris G, Leser R, Horsak B, Kornfeind P, Heller M, Baca A. Accuracy of the LPM tracking system considering dynamic position changes. J Sports Sci 2012;30(14):1503-1511. [doi: 10.1080/02640414.2012.712712] [Medline: 22906154]

175. Sathyan T, Shuttleworth R, Hedley M, Davids K. Validity and reliability of a radio positioning system for tracking athletes in indoor and outdoor team sports. Behav Res Methods 2012 Dec;44(4):1108-1114. [doi: 10.3758/s13428-012-0192-2] [Medline: 22477436]

176. Stelzer A, Fischer A, Vossiek M. A new technology for precise local position measurement-LPM. In: Proceedings of the IEEE MTT-S International Microwave Symposium Digest. Washington, DC: IEEE; 2004 Presented at: IEEE MTT-S International Microwave Symposium Digest; June 6-11, 2004; Fort Worth, TX p. 655-658. [doi: 10.1109/MWSYM.2004.1336072]

177. Stelzer A, Pourvoyeur K, Fischer A. Concept and Application of LPM- A Novel 3-D Local Position Measurement System. IEEE Trans Microw Theory Tech 2004 Dec;52(12):2664-2669. [doi: 10.1109/TMTT.2004.838281]

178. Pei L, Guinness R, Chen R, Liu J, Kuusniemi H, Chen Y, et al. Human behavior cognition using smartphone sensors. Sensors (Basel) 2013;13(2):1402-1424 [FREE Full text] [doi: 10.3390/s130201402] [Medline: 23348030]

179. Stratton G, Murphy R, Rosenberg M, Fergus P, Attwood A. Creating intelligent environments to monitor and manipulate physical activity and sedentary behavior in public health and clinical settings. In: Proceedings of the IEEE International Conference on Communications. Washington, DC: IEEE; 2012 Presented at: IEEE International Conference on Communications; June 10-15, 2012; Ottawa, ON p. 6111-6115. [doi: 10.1109/ICC.2012.6364974]

180. Stroiescu F, Daly K, Kuris B. Event detection in an assisted living environment. In: Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). Washington, DC: IEEE; 2011 Presented at: Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC); August 30-September 3, 2011; Boston, MA p. 7581-7584. [doi: 10.1109/IEMBS.2011.6091869]

181. Pei L, Chen R, Chen Y, Leppakoski H, Perttula A. Indoor/outdoor seamless positioning technologies integrated on smart phone. In: Proceedings of the International Conference on Advances in Satellite and Space Communications. Washington, DC: IEEE; 2009 Presented at: International Conference on Advances in Satellite and Space Communications; July 20-25, 2009; Colmar, France p. 141-145. [doi: 10.1109/SPACOMM.2009.12]

182. D'Souza M, Wark T, Karunanithi M, Ros M. Evaluation of realtime people tracking for indoor environments using ubiquitous motion sensors and limited wireless network infrastructure. Pervasive Mob Comput 2013 Aug;9(4):498-515. [doi: 10.1016/j.pmcj.2012.03.007]

183. Villarrubia G, De PJ, Bajo J, Corchado J. Real time positioning system using different sensors. In: Proceedings of the International Conference on Information Fusion. Washington, DC: IEEE; 2013 Presented at: International Conference on Information Fusion; July 9-12, 2013; Istanbul, Turkey p. 604-609.

184. Dunton GF, Whalen CK, Jamner LD, Floro JN. Mapping the social and physical contexts of physical activity across adolescence using ecological momentary assessment. Ann Behav Med 2007 Oct;34(2):144-153. [doi: 10.1080/08836610701566803] [Medline: 17927553]

185. Dunton GF, Liao Y, Intille S, Wolch J, Pentz MA. Physical and social contextual influences on children's leisure-time physical activity: an ecological momentary assessment study. J Phys Act Health 2011 Jan;8 Suppl 1:S103-S108. [Medline: 21350250]

186. Dunton GF, Kawabata K, Intille S, Wolch J, Pentz MA. Assessing the social and physical contexts of children's leisure-time physical activity: an ecological momentary assessment study. Am J Health Promot 2012 Feb;26(3):135-142. [doi: 10.4278/ajhp.100211-QUAN-43] [Medline: 22208410]

187. Dunton GF, Intille SS, Wolch J, Pentz MA. Investigating the impact of a smart growth community on the contexts of children's physical activity using Ecological Momentary Assessment. Health Place 2012 Jan;18(1):76-84 [FREE Full text] [doi: 10.1016/j.healthplace.2011.07.007] [Medline: 22243909]

188. Liao Y, Intille S, Wolch J, Pentz MA, Dunton GF. Understanding the physical and social contexts of children's nonschool sedentary behavior: an ecological momentary assessment study. J Phys Act Health 2014 Mar;11(3):588-595. [doi: 10.1123/jpah.2011-0363] [Medline: 23493261]

189. Spook JE, Paulussen T, Kok G, Van Empelen P. Monitoring dietary intake and physical activity electronically: feasibility, usability, and ecological validity of a mobile-based Ecological Momentary Assessment tool. J Med Internet Res 2013;15(9):e214 [FREE Full text] [doi: 10.2196/jmir.2617] [Medline: 24067298]

190. Liao Y, Intille SS, Dunton GF. Using ecological momentary assessment to understand where and with whom adults' physical and sedentary activity occur. Int J Behav Med 2015 Feb;22(1):51-61. [doi: 10.1007/s12529-014-9400-z] [Medline: 24639067]

191. Engelen L, Bundy AC, Lau J, Naughton G, Wyver S, Bauman A, et al. Understanding patterns of young children's physical activity after school - it's all about context: a cross-sectional study. J Phys Act Health 2015 Mar;12(3):335-339. [doi: 10.1123/jpah.2013-0153] [Medline: 24828415]

192. Greiner C, Makimoto K, Suzuki M, Yamakawa M, Ashida N. Feasibility study of the integrated circuit tag monitoring system for dementia residents in Japan. Am J Alzheimers Dis Other Demen 2007 May;22(2):129-136. [doi: 10.1177/1533317507299414] [Medline: 17545140]

193. Yamakawa M, Suto S, Shigenobu K, Kunimoto K, Makimoto K. Comparing dementia patients' nighttime objective movement indicators with staff observations. Psychogeriatrics 2012 Mar;12(1):18-26. [doi: 10.1111/j.1479-8301.2011.00380.x] [Medline: 22416825]

194. Hay S, Rassia S, Beresford A. Estimating personal energy expenditure with location data. In: Proceedings of the 8th IEEE International Conference on Pervasive Computing and Communications Workshops (PERCOM Workshops). Washington, DC: IEEE; 2010 Presented at: 8th IEEE International Conference on Pervasive Computing and Communications Workshops (PERCOM Workshops); March 29-April 2, 2010; Mannheim, Germany p. 304-309. [doi: 10.1109/PERCOMW.2010.5470650]

195. Ramulu PY, Chan ES, Loyd TL, Ferrucci L, Friedman DS. Comparison of home and away-from-home physical activity using accelerometers and cellular network-based tracking devices. J Phys Act Health 2012 Aug;9(6):809-817. [Medline: 21952052]

196. Kelly D, McLoone S, Logan B, Dishongh T. Single access point localisation for wearable wireless sensors. In: Proceedings of the 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society. Washington, DC: IEEE; 2008 Presented at: 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society; August 20-25, 2008; Vancouver, BC p. 4443-4446. [doi: 10.1109/IEMBS.2008.4650197]

197. Ben-Harush O, Carroll J, Marsh B. Using mobile social media and GIS in health and place research. Continuum: Journal of Media and Cultural Studies 2012 Oct;26(5):715-730. [doi: 10.1080/10304312.2012.706460]

198. Yeh S, Chang K, Wu C, Chu H, Hsu JY. GETA sandals: a footstep location tracking system. Pers Ubiquit Comput 2007 Feb 1;11(6):451-463. [doi: 10.1007/s00779-006-0098-z]

199. Shahid B, Kannan AA, Lovell NH, Redmond SJ. Ultrasound user-identification for wireless sensor networks. In: Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). Washington, DC: IEEE; 2010 Presented at: Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC); August 31-September 4, 2010; Buenos Aires, Argentina p. 5756-5759. [doi: 10.1109/IEMBS.2010.5627849]

200. Lan K, Shih W. On calibrating the sensor errors of a PDR-based indoor localization system. Sensors (Basel) 2013;13(4):4781-4810 [FREE Full text] [doi: 10.3390/s130404781] [Medline: 23575036]

201. Autographer. URL: http://www.autographer.com/ [accessed 2015-05-04] [WebCite Cache ID 6YGw4cjlu]

202. Narrative. URL: http://getnarrative.com/narrative-clip-1 [accessed 2015-05-04] [WebCite Cache ID 6YGwHeqDX]

203. MeCam. URL: http://www.mecam.me/collections/all-products/products/mecam?variant=388052188 [accessed 2015-05-04] [WebCite Cache ID 6YGwT4kCX]

204. uCorder. URL: http://www.ucorder.com/IRDC260r-pocket-camcorder.html [accessed 2015-05-04] [WebCite Cache ID 6YGwZxC9p]

205. Parashoot. URL: http://theparashoot.com/ [accessed 2015-05-04] [WebCite Cache ID 6YGwmeZsC]
206. Spy Emporium. URL: http://www.spyemporium.com/covert-wearable-cameras/5mp-hd-hidden-dvr-spy-camera-eyeglasses/ [accessed 2015-05-04] [WebCite Cache ID 6YGwqU2Eo]
207. VIEVU. URL: http://www.vievu.com/vievu-products/hardware/ [accessed 2015-05-04] [WebCite Cache ID 6YGwuJ9xy]
208. Panasonic. URL: http://security.panasonic.com/pss/security/products/ind/TW310/index.html [accessed 2015-05-04] [WebCite Cache ID 6YGx2RSpJ]
209. meMINI. URL: http://memini.com/ [accessed 2015-05-04] [WebCite Cache ID 6YGxHzYpt]
210. Pivothead. URL: http://www.pivothead.com/technology/originals/ [accessed 2015-05-04] [WebCite Cache ID 6YGxMEy7k]
211. Nixie. URL: http://flynixie.com/ [accessed 2015-05-04] [WebCite Cache ID 6YGxSVJZ8]
212. CA7CH. URL: http://www.ca7ch.com/preorder/ [accessed 2015-05-04] [WebCite Cache ID 6YGxXMpKU]
213. ELMO USA. URL: http://www.elmousa.com/qbic/wearable-cameras [accessed 2015-05-04] [WebCite Cache ID 6YGxcRGyH]
214. Vidcie. URL: http://www.vidcie.com/vidcie-cameras-kits/ [accessed 2015-05-04] [WebCite Cache ID 6YGxh08gG]
215. Ekahau. URL: http://www.ekahau.com/real-time-location-system [accessed 2015-05-14] [WebCite Cache ID 6YWEhsjm8]
216. Ubisense. URL: http://www.ubisense.net/en/products/smart-factory [accessed 2015-05-27] [WebCite Cache ID 6YqKS9lbd]
217. AeroScout. URL: http://www.aeroscout.com/t2-tag-family [accessed 2015-05-04] [WebCite Cache ID 6YGyOOGFg]
218. Zebra. URL: https://www.zebra.com/us/en/products/location-solutions/wherenet/wheretag.html [accessed 2015-05-04] [WebCite Cache ID 6YGyo0V4C]
219. Elpas. URL: http://www.elpas.com/Products/Healthcare-Asset-Tracking-Tag.aspx [accessed 2015-05-04] [WebCite Cache ID 6YGyyaIyb]
220. CenTrak. URL: http://www.centrak.com/products/ [accessed 2015-05-04] [WebCite Cache ID 6YGzfbSQ0]
221. TeleTracking. URL: http://www.teletracking.com/ [accessed 2015-05-04] [WebCite Cache ID 6YGzka6B8]
222. Sonitor. URL: http://www.sonitor.com/ [accessed 2015-05-04] [WebCite Cache ID 6YGzrAmB0]
223. Versus. URL: http://www.versustech.com/ [accessed 2015-05-04] [WebCite Cache ID 6YGzv8Cjj]
224. Radianse. URL: http://www.radianse.com/ [accessed 2015-05-27] [WebCite Cache ID 6YqKVF6Sd]
225. Secure Care. URL: http://www.securecare.com/healthcare-rtls [accessed 2015-05-27] [WebCite Cache ID 6YqKlYKwx]
226. Mojix. URL: http://www.mojix.com/products/elocation.php [accessed 2015-05-04] [WebCite Cache ID 6YH07G0z2]
227. AssetWorks. URL: http://www.assetworks.com/ [accessed 2015-05-04] [WebCite Cache ID 6YH0BLo9Y]
228. Tempsys. URL: http://www.tempsys.net/real-time-location-systems/ [accessed 2015-05-04] [WebCite Cache ID 6YH0EJ794]
229. Awarepoint. URL: http://www.awarepoint.com/ [accessed 2015-05-04] [WebCite Cache ID 6YH0HEO15]
230. Comita. URL: http://comita.net/en/comprehensive-security-solution/comita-rtls/ [accessed 2015-05-04] [WebCite Cache ID 6YH0LCdBK]
231. TrackIT. URL: http://www.thetrackit.com/products.php [accessed 2015-05-04] [WebCite Cache ID 6YH0O4WII]
232. Nebusens. URL: http://www.nebusens.com/index.php/en/products/n-core/hardware [accessed 2015-05-04] [WebCite Cache ID 6YH0S9BoJ]
233. Essensium. URL: http://www.essensium.com/ [accessed 2015-05-04] [WebCite Cache ID 6YH0UQ99f]
234. PLUS Location. URL: http://pluslocation.com/ [accessed 2015-06-24] [WebCite Cache ID 6ZWh66UWB]
235. Technical Life Care. URL: http://www.technicallifecare.com/RTLS-services.php [accessed 2015-05-04] [WebCite Cache ID 6YH0hI7eT]
236. AiRISTA. URL: http://www.airista.com/ [accessed 2015-05-14] [WebCite Cache ID 6YWGRcRw3]
237. Conduco. URL: http://www.conduco.co.uk/solutions/rtls-real-time-locating-systems [accessed 2015-05-04] [WebCite Cache ID 6YH10AGdB]
238. Luminosity. URL: http://www.luminosityhealth.com/index.php/luminosity-features/rtls [accessed 2015-05-04] [WebCite Cache ID 6YH12lYAG]
239. Purelink. URL: http://www.purelink.ca/fr/products/equipment-tracking-tag.php [accessed 2015-05-04] [WebCite Cache ID 6YH1652zg]
240. Sanitag. URL: http://www.sanitag.com/ [accessed 2015-05-04] [WebCite Cache ID 6YH190dzG]
241. AidA. URL: http://www.aidarfid.com/index.php/products-solutions/aida-rtls [accessed 2015-05-04] [WebCite Cache ID 6YH1ESOM5]
242. OpenRTLS. URL: http://www.openrtls.com/shop/product/tag-23 [accessed 2015-05-04] [WebCite Cache ID 6YH1G73NI]
243. BeSpoon. URL: http://bespoon.com/3d-precise-location-rtls/ [accessed 2015-05-04] [WebCite Cache ID 6YH1IRYuV]
244. Ecived. URL: http://www.ecived.com/en/product_show.aspx?id=18 [accessed 2015-05-04] [WebCite Cache ID 6YH1MZjK8]
245. Skytron. URL: http://www.skytron.us/products/pages/gen2ir.html [accessed 2015-05-04] [WebCite Cache ID 6YH1PoJOp]
246. LogiTag. URL: http://logi-tag.com/products/rtls/patientflow.html [accessed 2015-05-04] [WebCite Cache ID 6YH1RZ6GG]
247. RedPoint Positioning. Redpoint URL: http://www.redpointpositioning.com/products-services/?prod=radionodes [accessed 2015-05-04] [WebCite Cache ID 6YH1TCh6Q]
248. Borda Technology. URL: http://www.bordatech.com/products_hardware.html [accessed 2015-05-04] [WebCite Cache ID 6YH1UarcJ]
249. PointRF. URL: http://pointrf.com/technology.php [accessed 2015-05-04] [WebCite Cache ID 6YH1XpZPC]

250. Trackstick. URL: http://www.trackstick.com/products/mini/index.html [accessed 2015-05-04] [WebCite Cache ID 6YH28d2Em]

251. Trackershop. URL: http://www.trackershop-uk.com/_shop/personal-gps-trackers/pro-pod-5--personal-gps-tracker/ [accessed 2015-05-04] [WebCite Cache ID 6YH2InIKQ]

252. Gotek7. URL: http://www.gotek7.com/ [accessed 2015-05-27] [WebCite Cache ID 6YqLQpKwX]

253. CareWhere. URL: http://carewhere.co.uk/ [accessed 2015-05-04] [WebCite Cache ID 6YH2RrL9d]

254. PocketFinder. URL: http://www.pocketfinder.com/ [accessed 2015-05-04] [WebCite Cache ID 6YH2TAkRl]

255. Bluetrack. URL: http://bluetrackgpstrackers.co.uk/shop/gps-trackers/prime-lite/ [accessed 2015-05-04] [WebCite Cache ID 6YH2Un4ct]

256. Trackinapack. URL: http://www.trackinapack.co.uk/advanced-gps-tracking-device [accessed 2015-05-05] [WebCite Cache ID 6YIMECB6f]

257. Protect my kids. URL: http://www.protectmykids.co.uk/gps-tracking-device-kids-monitoring-protect-kids/ [accessed 2015-05-05] [WebCite Cache ID 6YIMJii1w]

258. Amber Alert GPS. URL: http://www.amberalertgps.com/ [accessed 2015-05-05] [WebCite Cache ID 6YIMM4O48]

259. TracLogik. URL: http://www.traclogik.co.uk/product/covert-2000/ [accessed 2015-05-05] [WebCite Cache ID 6YIMXShO8]

260. Laipac. URL: http://www.laipac.com/ [accessed 2015-05-27] [WebCite Cache ID 6YqLl5Lgd]

261. Loc8tor. URL: http://www.loc8tor.com/ [accessed 2015-05-05] [WebCite Cache ID 6YIMp8VEy]

262. Meitrack. URL: http://www.meitrack.net/portable-gps-tracker/ [accessed 2015-05-05] [WebCite Cache ID 6YIMrHrXm]

263. SonikGPS. URL: http://sonikgps.com/sonik_personal_tracker.html [accessed 2015-05-05] [WebCite Cache ID 6YIMvH9kC]

264. Global Tracking Group. URL: http://www.globaltrackinggroup.com/gps-tracking-device-ubi5000e.cfm [accessed 2015-05-05] [WebCite Cache ID 6YIMxKCgw]

265. GPSintegrated. URL: http://gpsintegrated.com/products/personal-tracker/handheld-gpsgsm-tracker-gh1202 [accessed 2015-05-05] [WebCite Cache ID 6YIN0tHWR]

266. buddi. URL: https://www.buddi.co.uk/ [accessed 2015-05-05] [WebCite Cache ID 6YIN4sVi3]

267. Keytracker. URL: http://www.keytracker.com/products_personnelgps.php [accessed 2015-05-05] [WebCite Cache ID 6YIN7lmLp]

268. RM Tracking. URL: http://www.rmtracking.com/ [accessed 2015-05-05] [WebCite Cache ID 6YINCZiYk]

269. LandAirSea. URL: https://www.landairsea.com/gps-tracker [accessed 2015-05-05] [WebCite Cache ID 6YINH1Iyz]

270. Dynaspy. URL: https://www.dynaspy.com/gps-tracking-devices/real-time-gps-trackers/worldtracker-enduro-pro-gps-tracker [accessed 2015-05-05] [WebCite Cache ID 6YINKLc1w]

271. WheribleGPS. URL: http://www.wheriblegps.com/products/wheritrack-device/ [accessed 2015-05-05] [WebCite Cache ID 6YIOK1FGs]

272. iLOC. URL: https://www.iloctech.com/triloc/purchase [accessed 2015-05-05] [WebCite Cache ID 6YIOLnr4b]

273. GTX Corp. URL: http://gtxcorp.com/ [accessed 2015-05-27] [WebCite Cache ID 6YqM3nSni]

274. BioSensics. URL: http://www.biosensics.com/shop/pamsys-motion-sensor-with-gps/ [accessed 2015-05-05] [WebCite Cache ID 6YIOY4x8d]

275. Recon Instruments. URL: http://www.reconinstruments.com/products/jet/ [accessed 2015-05-05] [WebCite Cache ID 6YIOa78Hb]

276. Nike. URL: https://secure-nikeplus.nike.com/plus/products/sport_watch/ [accessed 2015-05-05] [WebCite Cache ID 6YIOcS6qr]

277. Garmin. URL: http://www.garmin.com/en-GB [accessed 2015-05-14] [WebCite Cache ID 6YWKYQ9Ns]

278. Revolutionary Tracker. URL: http://www.revolutionarytracker.com/pro/ [accessed 2015-05-05] [WebCite Cache ID 6YIPXPyAz]

279. Everon. URL: http://www.everon.fi/en/solutions/vega-gps-safety-solution-and-bracelet [accessed 2015-05-05] [WebCite Cache ID 6YIPZia4b]

280. Trax. URL: https://traxfamily.zendesk.com/hc/en-gb/articles/200310982-Tracker-specifications [accessed 2015-05-05] [WebCite Cache ID 6YIQ5NeF1]

281. Ninja Tracking. URL: http://www.ninjatracking.co.uk/prime-at-lite [accessed 2015-05-05] [WebCite Cache ID 6YIQ6mio2]

282. Personal GPS Trackers. URL: http://www.personalgpstrackers.co.uk [accessed 2015-05-05] [WebCite Cache ID 6YIQ988PK]

283. Retrievor. URL: http://www.retrievor.com/pre-order [accessed 2015-05-05] [WebCite Cache ID 6YIQFD6pa]

284. Duotraq. URL: http://www.duotraq.com/ [accessed 2015-05-27] [WebCite Cache ID 6YqMBk0Vz]

285. Mindme. URL: http://www.mindme.care/ [accessed 2015-05-05] [WebCite Cache ID 6YIQIY6oY]

286. Bubble Tracker. URL: http://www.bubbletracker.com/ [accessed 2015-05-05] [WebCite Cache ID 6YIQKMiyU]

287. Maddison R, Ni Mhurchu C. Global positioning system: a new opportunity in physical activity measurement. Int J Behav Nutr Phys Act 2009;6:73 [FREE Full text] [doi: 10.1186/1479-5868-6-73] [Medline: 19887012]

288. Krenn PJ, Titze S, Oja P, Jones A, Ogilvie D. Use of global positioning systems to study physical activity and the environment: a systematic review. Am J Prev Med 2011 Nov;41(5):508-515 [FREE Full text] [doi: 10.1016/j.amepre.2011.06.046] [Medline: 22011423]

289. Kamel Boulos MN, Berry G. Real-time locating systems (RTLS) in healthcare: a condensed primer. Int J Health Geogr 2012;11:25 [FREE Full text] [doi: 10.1186/1476-072X-11-25] [Medline: 22741760]

290. Stojanović D, Stojanović N. Facta Universitatis, Series: Automatic Control and Robotics. 2014. Indoor localization and tracking: methods, technologies and research challenges URL: http://casopisi.junis.ni.ac.rs/index.php/FUAutContRob/article/download/208/88 [accessed 2015-07-31] [WebCite Cache ID 6aRElrDiw]

291. Liu H, Darabi H, Banerjee P, Liu J. Survey of Wireless Indoor Positioning Techniques and Systems. IEEE Trans Syst Man Cybern C Appl Rev 2007 Nov;37(6):1067-1080. [doi: 10.1109/TSMCC.2007.905750]

292. Curran K, Furey E, Lunney T, Santos J, Woods D, McCaughey A. An evaluation of indoor location determination technologies. Journal of Location Based Services 2011 Jun;5(2):61-78. [doi: 10.1080/17489725.2011.562927]

293. Koyuncu H, Yang S. A survey of indoor positioning and object locating systems. International Journal of Computer Science and Network Security 2010;10(5):121-128.

294. Kelly P, Doherty A, Berry E, Hodges S, Batterham AM, Foster C. Can we use digital life-log images to investigate active and sedentary travel behaviour? Results from a pilot study. Int J Behav Nutr Phys Act 2011;8:44 [FREE Full text] [doi: 10.1186/1479-5868-8-44] [Medline: 21599935]

295. Macdonald-Wallis K, Jago R, Sterne JA. Social network analysis of childhood and youth physical activity: a systematic review. Am J Prev Med 2012 Dec;43(6):636-642. [doi: 10.1016/j.amepre.2012.08.021] [Medline: 23159259]

296. Kelly P, Marshall SJ, Badland H, Kerr J, Oliver M, Doherty AR, et al. An ethical framework for automated, wearable cameras in health behavior research. Am J Prev Med 2013 Mar;44(3):314-319. [doi: 10.1016/j.amepre.2012.11.006] [Medline: 23415131]

Abbreviations

AOA: angle of arrival
BL: battery life
BLE: Bluetooth low energy
CCTV: closed-circuit television
CR: camera resolution
Dim: dimensions
EMA: ecological momentary assessment
FPS: frames per second
GIS: geographic information system
GPS: global positioning system
IC: integrated circuit
IEEE: Institute of Electrical and Electronics Engineers
I/O: indoor/outdoor
IR: infrared
Man: manufacturer
MET: metabolic equivalent
MVPA: moderate-to-vigorous physical activity
N/A: not applicable
NIHR CLAHRC for EM: National Institute for Health Research Collaboration for Leadership in Applied Health Research and Care—East Midlands
PALMS: Personal Activity and Location Measurement System
Refs: references
RF: radio frequency
RFID: radio-frequency identification
RSSI: received signal strength indicator
RTLS: real-time locating system
SF: sampling frequency
TDOA: time difference of arrival
TOA: time of arrival
TOF: time of flight
TWR: two-way ranging
UWB: ultra wide band

Original Paper

"If I Were Nick": Men's Responses to an Interactive Video Drama Series to Support Smoking Cessation

Joan L Bottorff[1,2*], RN, FCAHS, FAAN, PhD; Gayl Sarbit[3*], PhD; John L Oliffe[4*], RN, PhD; Mary T Kelly[3*], MA; Maria Lohan[5*], PhD; Sean Stolp[3*], MA; Paul Sharp[3*], MSc

[1]School of Nursing, Faculty of Health and Social Development, University of British Columbia, Kelowna, BC, Canada

[2]Faculty of Health Sciences, Australian Catholic University, Melbourne, Australia

[3]Institute for Healthy Living and Chronic Disease Prevention, University of British Columbia, Kelowna, BC, Canada

[4]School of Nursing, Faculty of Applied Science, University of British Columbia, Vancouver, BC, Canada

[5]School of Nursing and Midwifery, Queen's University Belfast, Belfast, Ireland

[*]all authors contributed equally

Corresponding Author:
Joan L Bottorff, RN, FCAHS, FAAN, PhD
School of Nursing
Faculty of Health and Social Development
University of British Columbia
1147 University Road, Art 223
Kelowna, BC, V1V 1V7
Canada
Phone: 1 2508078627
Fax: 1 2508078629
Email: joan.bottorff@ubc.ca

Abstract

Background: Men continue to smoke in greater numbers than women; however, few interventions have been developed and tested to support men's cessation. Men tend to rely on quitting strategies associated with stereotypical manliness, such as willpower, stoicism, and independence, but they may lack the self-efficacy skills required to sustain a quit. In this paper, we describe the development of and reception to an interactive video drama (IVD) series, composed of 7 brief scenarios, to support and strengthen men's smoking cessation efforts. The value of IVD in health promotion is predicated on the evidence that viewers engage with the material when they are presented characters with whom they can personally identify. The video dramatizes the challenges unfolding in the life of the main character, Nick, on the first day of his quit and models the skills necessary to embark upon a sustainable quit.

Objective: The objective was to describe men's responses to the *If I were Nick* IVD series as part of a study of QuitNow Men, an innovative smoking cessation website designed for men. Specific objectives were to explore the resonance of the main character of the IVD series with end-users and explore men's perceptions of the effectiveness of the IVD series for supporting their quit self-management.

Methods: Seven brief IVD scenarios were developed, filmed with a professional actor, and uploaded to a new online smoking cessation website, QuitNow Men. A sample of 117 men who smoked were recruited into the study and provided baseline data prior to access to the QuitNow Men website for a 6-month period. During this time, 47 men chose to view the IVDs. Their responses to questions about the IVDs were collected in online surveys at 3-month and 6-month time points and analyzed using descriptive statistics.

Results: The majority of participants indicated they related to the main character, Nick. Participants who "strongly agreed" they could relate to Nick perceived significantly higher levels of support from the IVDs than the "neutral" and "disagree" groups ($P<.001$, $d=2.0$, $P<.001$, $d=3.1$). The "agree" and "neutral" groups were significantly higher on rated support from the videos than the "disagree" ($P<.001$, $d=2.2$, $P=.01$, $d=1.5$). Participants' perception of the main character was independent of participant age, education attainment, or previous quit attempts.

Conclusions: The findings suggest that IVD interventions may be an important addition to men's smoking cessation programs. Given that the use of IVD scenarios in health promotion is in its infancy, the positive outcomes from this study signal the potential for IVD and warrant ongoing evaluation in smoking cessation and, more generally, men's health promotion.

Introduction

Background

Health promotion and tobacco control efforts have resulted in a steady decline in the prevalence of cigarette smoking in Canada and the United States over the past 50 years; however, the highest rates of tobacco use continue to be among adult men, which far exceed population averages [1]. To address the high prevalence of smoking among men, novel gender-specific resources are needed [2,3].

Scholars in men's health have argued that men prefer to self-manage their health, but at the same time, their health behaviors are strongly influenced by socially produced constructs of masculinity [4,5]. In terms of smoking cessation, men tend to rely on quitting strategies associated with stereotypical manliness, such as willpower, stoicism, and independence [6]. Although men often rely on willpower and a "cold turkey" approach to quit smoking [7], it is well established that this is one of the least effective means of smoking cessation [8]. Despite the availability of a range of smoking cessation programs, there are few programs tailored specifically for men [3]. However, emerging evidence related to men's experiences and preferences with smoking cessation, along with established approaches to supporting behavior change, provide an important foundation for tailored interventions. For example, men's preferences for self-managing their quit and making their own decisions about when and how to quit have been reported [9]. There is evidence that these gendered preferences can be supported and enhanced by the use of online information and communication technologies in men's health promotion [10]. In addition, men's interest in learning about others' real-life cessation efforts as a way to develop effective plans for quitting and acquiring strategies that they might also use has been described [2]. Information to support anticipatory planning and skill building can raise self-efficacy and increase confidence and the ability to stop smoking. There is agreement that strengthening self-efficacy by garnering skills and confidence to respond to challenging situations, stressful events, and cravings is an important component in providing support for smoking cessation [11]. Tailoring supports for men's anticipatory planning and skill building related to quitting could, therefore, provide important additions to cessation programs.

Interactive video drama (IVD) is an emergent media tool with potential for the field of smoking cessation interventions for men. The value of IVD is predicated on the evidence that viewers engage with, and are more influenced by, material when they are presented with a character with whom they personally identify. In a feasibility trial conducted in Northern Ireland, Lohan et al [12] developed and are testing a 4-week IVD program for sex education classes in 7 schools with the aim of reducing unintended teenage pregnancy. Based on observational data of over 700 young men, their video intervention is theorized to strengthen the intentions of male teens to avoid unplanned pregnancy by influencing the psychosocial variables involved in sexual behaviors and choices [12]. To date, IVDs have not been utilized or evaluated in smoking cessation programs.

The capacity for IVDs to incorporate multiple features that are important to men, and at the same time, promote anticipatory thinking and model skills associated with strong self-efficacy makes the approach a good fit to support men's cessation efforts. The drama-based approach, characteristic of IVDs, offers an important way to engage men and maintain their interest by using men's authentic voices to present challenging situations wherein the main character models change and effective problem solving– such as the decision to stop smoking and maintain a quit. IVDs also may be effective for men because of the motivational appeal of interactive resources, the popularity of videos with men, the high level of interest in learning about others' "real life" experiences, and the desire by men to see where they stand in relation to choices made by other men.

As part of a study to design and evaluate an online smoking cessation resource designed specifically for men, QuitNow Men [13], we developed an IVD series, *If I were Nick*, as a resource to support men's cessation efforts. This IVD series was included on the QuitNow Men website along with a broad array of interactive tools and tips for men interested in quitting smoking, such as a questionnaire to determine their level of nicotine dependence, a calculator to show the financial cost of smoking, expert information and motivational videos, strategies for quitting, and quizzes to test tobacco use knowledge. The IVD provided the opportunity to influence men's beliefs about quitting through their identification with the central character and dramatized evidenced-based content to strengthen smoking cessation self-efficacy by prompting reflection on smoking triggers, encouraging planning and response strategies for stressors before they arise, and modeling how to incorporate these strategies into daily life to maximize a successful quit. In this paper, we describe the development of this IVD series and report findings related to the resonance of the main character of the IVD series with end-users, and men's perceptions of the effectiveness of the IVDs for supporting their quit self-management.

The Framework and Development of If I Were Nick

The *If I Were Nick* videos were developed in accordance with the theoretical framework described in the work of Lohan et al [12] and Aventin et al [14] that blends social theory on gender norms and masculinity with cognitive learning theory and the

theory of planned behavior, which have demonstrated that attitudes, perceived control, social context, and intentions influence behavior change [15,16]. A script was written and a story board developed by the team to illustrate high-risk situations that might happen to Nick, a construction worker, during the first day of a quit. Ideas for the high-risk situations were based on focus group data collected with men who smoke [2] and cues for smoking among male smokers that have been identified in the literature [17]. The role of Nick was performed by a professional actor who met the qualifications established by the research team to best engage the target audience: age 25-35 years, genuine and personable on camera, strong male voice, and good range of emotion. Seven scenarios were filmed, each scenario running about one minute. Following each

scenario, two online questions prompted viewers to reflect and respond: "How would you feel if you were Nick in this situation?" and "How would you act in this situation in order to remain smoke free?" Viewers were presented with a multiple choice list of 4-6 answer options and asked to click on their selection. The questions were designed to promote anticipatory thinking around smoking cessation and encourage men to consider how they might respond when facing the challenges of quitting. After making their selection, the viewer was able to see how other men responded. Table 1 summarizes the IVD scenarios and reflective questions. Table 2 provides additional details for one scenario from the IVD series, entitled "Morning Routine", including the response items for the self-reflective and action-oriented questions.

Table 1. Story development of the interactive video drama, *If I Were Nick*.

Video scenario	High-risk quit situation	Self-reflective question	Action question
Meet Nick (Multimedia Appendix 1)	At home, Nick tells us he is quitting smoking today and has tried many times in the past. He wants us to understand the experience through his eyes.	If I were Nick, how would I feel on the first day of my quit?	If I were Nick, what tactic would I choose?
Morning Routine (Multimedia Appendix 2)	A flashback in Nick's mind shows us how the usual morning coffee with a cigarette is the first challenge of the day.	If I were Nick, how would I feel about changing my morning routine?	If I were Nick, what would I change?
On the Road (Multimedia Appendix 3)	Nick pulls into traffic on the way to work and is forced to wait at a red light, which causes him to fidget and feel anxious.	If I were Nick, how would I feel about being in my car without any smokes?	If I were Nick, how would I deal with cravings while driving?
I Need a Break (Multimedia Appendix 4)	On the job site, Nick is conflicted about taking a break with his co-workers, because they all smoke. He worries that if he tells his foreman he's now a non-smoker, he won't get as many breaks.	If I were Nick, how would I feel about telling my boss and fellow workers that I've quit smoking?	If I were Nick, what could I do at work to avoid temptation?
Out with the Guys (Multimedia Appendix 5)	Nick meets friends at the bar to have a beer and watch the hockey game after work. He's concerned how his friends will respond to his decision to quit.	If I were Nick, what do I think my friends would say when I tell them I've quit smoking?	If I were Nick, how would I respond to pressure from my friends to smoke?
Stressed Out (Multimedia Appendix 6)	Nick enters a convenience store to buy milk and tries to navigate a stressful situation on the telephone while paying the cashier. The interaction rattles him and he notices cigarettes next to the checkout.	If I were Nick, how would I feel about not having a cigarette when I'm stressed?	If I were Nick, how would I stay smoke free in a stressful situation?
On Track (Multimedia Appendix 7)	Nick is lying on the couch thinking about the events of the day and what tomorrow will be like.	If I were Nick, how would I feel after my first quit day?	If I were Nick, how would I keep myself on track?

Table 2. Scenario #2 from *If I Were Nick*.

Video segment	High-risk situation	Reflective question	Action question
Morning Routine (Multimedia Appendix 2)	A flashback in Nick's mind shows us how the usual morning coffee with a cigarette is the first challenge of the day.	If I were Nick, how would I feel about changing my morning routine? (a) I'm proud that I'm making healthy choices for myself, (b) I'll miss some things I enjoy but I'll try new things, (c) I feel that I'm strong enough to give it a shot, (d) I'm on the fence about changing my routine, (e) I'm comfortable making changes to my normal routine.	If I were Nick, what would I change? (a) Change the order of my morning routine, (b) Try a different drink, brand of coffee or mug, (c) Move to a new place to drink coffee and eat breakfast, (d) Jump in the shower as soon as I wake up, (e) Do something active in the morning (walk the dog, lift weights at home, do push-ups, hit the gym).

XSL•FO
RenderX

Methods

Overview

The current study was conducted using an exploratory descriptive design and received approval from a university ethics review committee.

Sample

A convenience sample of men who smoked and were interested in quitting were recruited from across Canada using social media and online advertisements. A total of 117 men consented to participate, completed a baseline survey, and were provided access to an online smoking cessation resource, QuitNow Men [13]. As one of the resources on QuitNow Men [13], the streamed IVD series, *If I Were Nick*, was available under a tab labeled "Tactics & Tools", as one of several resources to support smoking cessation.

The men were invited to use the website independently over a period of 6 months and could choose to use any or all of the resources available. Of the 117 men, 75 completed at least one of the follow-up surveys (response rate=64.1%), and 47 (62.7%) reported that they had viewed one or more of the IVD scenarios. These 47 men provided the data for this study.

Data Collection

All data collection was completed using an online survey platform at three time points: baseline, 3 months, and 6 months. Participants were provided with two reminders to complete each follow-up survey.

At baseline, demographics were measured including age, marital status, education, and employment as well as participant smoking behaviors. Men's use of and responses to the IVDs were assessed in the follow-up surveys. A 3-point Likert scale question ("none", "some", "all") was used to assess how many of the Nick video scenarios participants had watched. Participants were asked to respond to the statement "Nick was someone I could relate to" using a 5-point Likert response scale ("strongly disagree" to "strongly agree"). Four 5-point Likert scale questions ("strongly disagree" to "strongly agree") were included to assess how well the videos supported participants'

self-management by providing them with cognitive and behavioral strategies. The questions were "The videos made me think about issues I hadn't thought about before", "As I watched Nick, I thought about how I needed to change my own smoking routines", "The videos helped me prepare ahead for situations where I might be tempted to smoke", and "The videos helped me realize that I can be as strong as Nick in staying quit". These four items were then summed to create an IVD support score, rated from 4-20, which measured the degree of support the participants derived from viewing the videos. To assess whether the 4 items that were totaled formed a reliable score, Cronbach alpha was computed. The alpha for the four items was .827, indicating that the items formed a scale with good internal consistency. Finally, a 1-item, 5-point Likert scale ("very satisfied" to "very dissatisfied") assessed participants' satisfaction with the overall website (ie, "Overall, how satisfied were you with the QuitNow website?").

Data Analyses

As men could complete a 3-month survey and a 6-month survey, a hierarchy was developed for data inclusion. If an individual increased the number of IVD scenarios they watched between the 3-month and 6-month time point (ie, "none" to "some," or "some" to "all"), 6-month data were used. If there was no increase between time points, 3-month data were used, as it was deemed the videos would be more easily recalled at that time. Descriptive statistics were used to analyze the participants' survey responses and analysis of variance (ANOVA) was used to examine the relationship between levels of relating to Nick and participants' perceived support from the IVDs. Spearman rho statistic was calculated to assess possible relationships between levels of relating to Nick and participants' overall website satisfaction. All analyses were conducted using SPSS version 22.

Results

At baseline, all participants were smokers who were interested in quitting. They smoked on average 14.2 cigarettes a day (SD 7.65) and had been smoking on average for 21.4 years (SD 11.4). Table 3 provides the demographic characteristics of the sample.

Table 3. Participant demographics.

	Mean (SD) or n (%)
Age	37.2 (10.3)
Smoking behavior	
Cigarettes smoked per day	14.2 (7.65)
Years smoking	21.4 (11.4)
Marital status	
Married or common law	25 (53%)
Single	16 (34%)
Divorced or separated	6 (13%)
Education	
Attended or completed university	27 (57%)
Completed high school	9 (19%)
Completed a non-university degree (technical, trade)	8 (17%)
Not completed high school	3 (6%)
Main activity	
Work for pay	34 (72%)
Unemployed	5 (11%)
Going to school	4 (9%)
Recovering from illness or disability	3 (6%)
Caring for family	1 (2%)

Use of Website and Exposure to Interactive Video Dramas

During the 6-month time frame, participants reported that they accessed the website at least once and on average "4-6 times". All of the IVDs were watched by 40% of these participants; the remainder stated they had watched "some" of the scenarios.

Responses to the Interactive Video Dramas

Responses to the statement, "Nick was someone I could relate to", indicated that the majority of participants related to the character of Nick (strongly agree=17%, agree=40%). None indicated they "strongly disagreed", and 6 men (13%) indicated they "disagreed" that Nick was someone with whom they could relate. The remaining men (30%) rated their relatability to Nick as "neutral". The levels of relating to Nick did not significantly differ in terms of participants' age (P=.47), number of previous quit attempts in the past two years (P=.19), or educational background (P=.83). However, the participants' level of relating to Nick was positively correlated with their overall website satisfaction (r_{s45}=.63, P<.001). The majority of participants, ranging from 52-72% of the men, "agreed" or "strongly agreed" with all four statements about IVD support. Responses to each statement are shown in Figure 1.

With a possible range of 4-20, the mean IVD support score was 10.5 (SD 2.89). A statistically significant difference was found among the four levels of relating to Nick compared with perceived support from the videos ($F_{3,43}$=15.181, P<.001). Figure 2 shows the mean IVD support score by level of relating to Nick. Post hoc Tukey HSD tests indicated that men who "strongly agreed" they could relate to Nick perceived significantly higher levels of support from the IVDs than those in the "neutral" and "disagree" groups (P<.001, d=2.0, P<.001, d=3.1) . Those in the "agree" and "neutral" groups were significantly higher on rated support from the videos than the "disagree" group (P<.001, d=2.2, P=.01, d=1.5).

Figure 1. Participant responses to statements regarding the interactive video dramas.
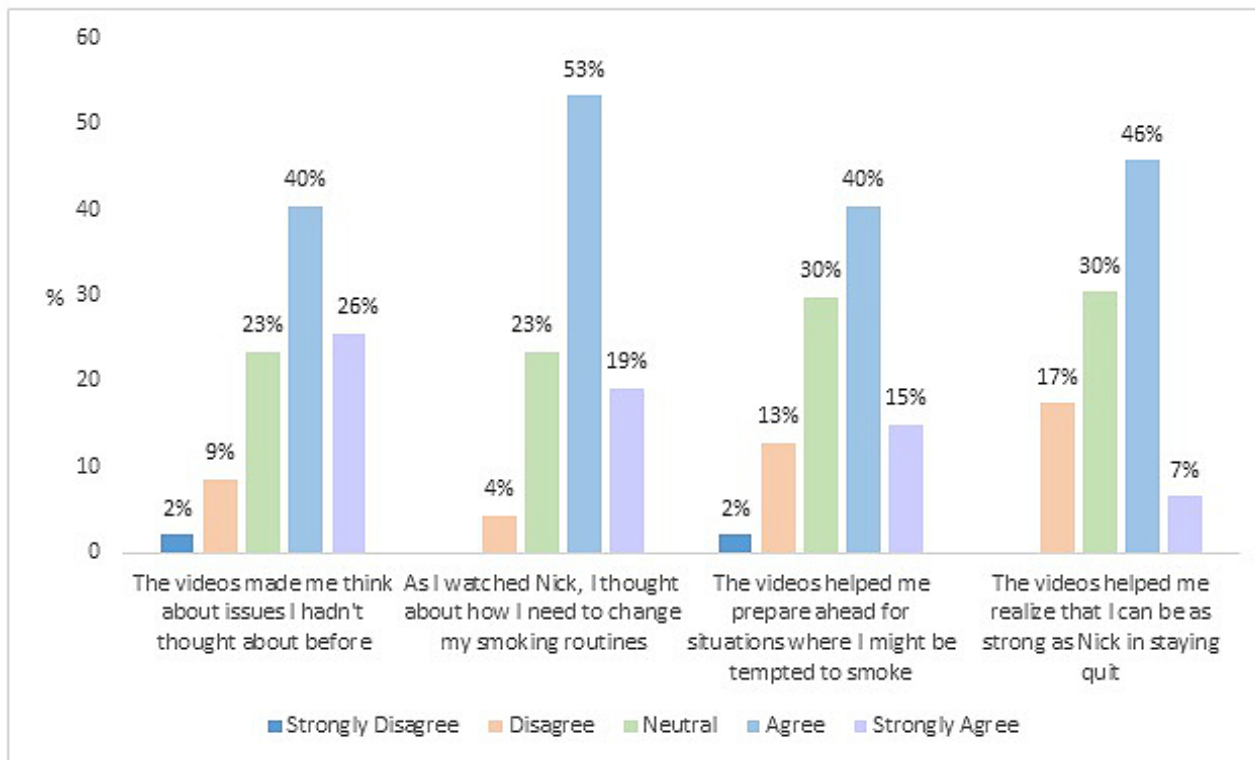
Figure 2. Means and standard error bars comparing relating to Nick groups on rated support from interactive video dramas.
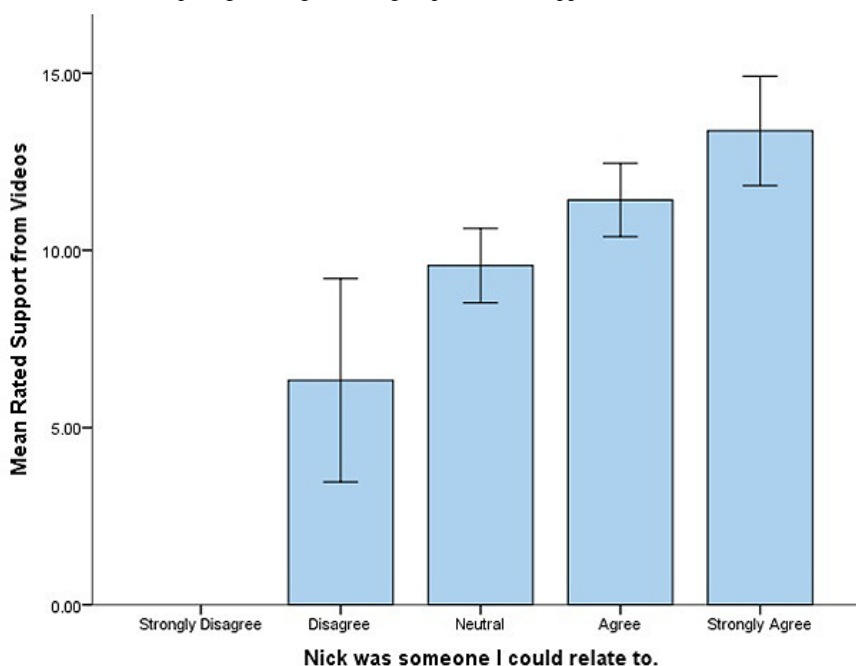
Discussion

Principal Findings

The current study focused on examining the IVDs as one component of a multipart gender-sensitive website designed to support men's smoking cessation. IVDs are an innovative approach in men's health promotion research [12], and the current study findings document the first application of IVDs in smoking cessation. More specifically, it is the first to target men who smoke. Results indicate that the IVDs included in the *If I were Nick* series were perceived as supportive by many men who were interested in reducing and quitting smoking.

The finding that only 47 of the 117 potential respondents viewed at least some of the IVDs may have been influenced by the placement of the IVDs on the QuitNow Men website [13]. The IVDs could not be directly accessed from the home page (*If I*

were Nick was accessible only after clicking on the tab "Tactics & Tools") and without a visual identifier on the home page may have been missed. However, it is encouraging that almost two-thirds of the men viewed at least some of the IVDs, and among those who viewed the IVDs, 40% viewed all 7 of the scenarios, providing an indication of the acceptability of this approach to men. The potential success of IVDs, as indicated by men's uptake of and perceived support from them, affirms the need for ongoing evaluation of the effectiveness of IVDs as they relate to smoking cessation and, more generally, men's health promotion.

Our finding that a majority of men related to Nick suggests that it may be important to ensure that the main IVD character is someone with whom the audience can identify. It appears, though, that the character may not need to perfectly mirror the demographics of the target audience. For instance, relatability was not associated with viewers' age or previous quit attempts. It was especially noteworthy that the character's employment as a construction worker did not mitigate identification with the character by middle class men working in the tertiary sector. These findings may suggest that men were connecting with the struggles of Nick and his reaction to the challenges that he faced, rather than demographic identity (eg, ethnicity, class, age, marital status). The modeling and permission of "other" men is also known to deeply influence the actions and non-actions of men in relation to their health [4]. Bearing both these points in mind, the resonance of Nick with end-users may flow from *how*, as well as *what* he does—rather than *who* he is perceived to be. It is possible that Nick's workmanlike routine and its segmented connections to smoking resonated with the participants' desire to feel more competent and prepared to execute a quit. As such, participants could have identified not only with Nick, but with his challenges around smoking cessation—amid being prompted to choose their own strategies. So while men's identity has often been linked to their jobs or careers, identification with Nick and the quitting challenges represented in the IVDs were less likely to depend on alignments to the end-users' employment status or specific job.

There were some men who did not relate with Nick. This could reflect a desire to "stand out from the herd", and dis-identify with this "ordinary" man, or perhaps claim identities that counter and contrast some aspect of Nick (masculine representation, age, class, occupation, culture). The failure to identify with Nick might also reflect participants' uncertainty about their readiness to quit—wherein they critiqued the IVD product or its utility rather than the process of quitting depicted therein. Future qualitative research might explore the variation in the men's identification with Nick, as well as variation in response to Nick based on age difference, given the participants in this study were on average 37 years old.

Although information on the challenges faced while quitting smoking are widely available on the Internet, the men's responses to *If I were Nick* point to the potential merit of the IVDs in presenting this information through peer dialogue and real-life situations. The IVDs were designed to promote the internal reflection of what men may think, feel, and do, in certain situations. Men often avoid discussing emotion as it is not considered "manly" or may be misconstrued as weakness. By providing the internal dialogue of a fellow man and his struggles, the IVDs are intended to normalize feelings that often go undiscussed among the male population. These emotions are further supported through seeing the responses of other men once they have identified and locked in their responses using the reflective questions.

The IVDs were also designed to empower the audience by assisting men in visualizing and preparing for the challenges of quitting and exposing them to the requisite skills needed to sustain their quit thereby boosting self-efficacy, a critical mediating influence on smoking cessation outcomes [18,19]. The findings from the current study indicated that preparing and planning strategies, and providing choices in the context of an IVD, hold potential for enabling men to action their interest in reducing or quitting smoking. The disaggregated day's events, so typical of many men's lives, may have prompted participants to deconstruct specific routines and rituals common to their everyday lives while prospectively planning what they might do in similar circumstances. In an online environment, a men's IVD smoking cessation intervention also allows for that planning and strategy work to take place in an anonymous setting, free of experts, which may be perceived by men as taking action rather than receiving help—an approach in accord with masculine values.

There is much uncertainty about best practices for IVDs in health promotion, and in the specific context of men's smoking cessation, clearly, there is much to learn. Beyond identifying (or not) with the central character (which could be remedied through technologies enabling end-users to select an avatar or choose from a range of characters or insert their own image to a vignette), the number and duration of the IVD scenarios might influence the uptake and "completion" of the series. Herein lies the need for future large, longitudinal, comparative research to better apprehend and predict what media and mechanisms might prevail within an ever changing "virtual" forum for men's health promotion. Future research might also measure quit rates to assess the impact of IVDs on men's smoking cessation.

Conclusions

IVDs are an exciting new approach to support men's smoking cessation that provide the opportunity to combine emotional, social, cognitive, and behavioral aspects of quitting. A well-developed IVD can incorporate the principles of men's health promotion, include multiple features that appeal to men and convey valuable smoking cessation tactics in an instructive yet empowering way. When men gain new information and believe they are skilled enough to apply it successfully, they are more likely to try and to succeed in quitting smoking. Given that the use of IVDs in health promotion is in its infancy, the positive outcomes from this study suggest the potential for IVDs and warrant moving forward in this area.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Meet Nick.

[MOV File, 67MB - jmir_v17i8e190_app1.mov]

Multimedia Appendix 2

Morning routine.

[MOV File, 46MB - jmir_v17i8e190_app2.mov]

Multimedia Appendix 3

On the road to work.

[MOV File, 58MB - jmir_v17i8e190_app3.mov]

Multimedia Appendix 4

Need a break.

[MOV File, 32MB - jmir_v17i8e190_app4.mov]

Multimedia Appendix 5

Out with the guys.

[MOV File, 38MB - jmir_v17i8e190_app5.mov]

Multimedia Appendix 6

Stressed out.

[MOV File, 50MB - jmir_v17i8e190_app6.mov]

Multimedia Appendix 7

On track.

[MOV File, 40MB - jmir_v17i8e190_app7.mov]

References

1. Health Canada. Canadian Tobacco Use Monitoring Survey (CTUMS) 2012. Ottawa: Health Canada; 2012. URL: http://www.hc-sc.gc.ca/hc-ps/tobac-tabac/research-recherche/stat/_ctums-esutc_2012/ann-eng.php [accessed 2015-03-26] [WebCite Cache ID 6XKCTRHev]

2. Bottorff J, Oliffe J, Sarbit G. Smoke free men: Competing and connecting to quit. Am J Health Promot 2015 (in review) 2015:- (forthcoming).

3. Okoli CT, Torchalla I, Oliffe JL, Bottorff JL. Men's smoking cessation interventions: a brief review. Journal of Men's Health 2011 May;8(2):100-108. [doi: 10.1016/j.jomh.2011.03.003]

4. Creighton G, Oliffe JL. Theorising masculinities and men's health: A brief history with a view to practice. Health Sociology Review 2014 Dec 17;19(4):409-418. [doi: 10.5172/hesr.2010.19.4.409]

5. Galdas PM, Cheater F, Marshall P. Men and health help-seeking behaviour: literature review. J Adv Nurs 2005 Mar;49(6):616-623. [doi: 10.1111/j.1365-2648.2004.03331.x] [Medline: 15737222]

6. White C, Oliffe JL, Bottorff JL. Tobacco and the invention of quitting: a history of gender, excess and will-power. Sociol Health Illn 2013 Jun;35(5):778-792 [FREE Full text] [doi: 10.1111/j.1467-9566.2012.01529.x] [Medline: 23600564]

7.   Bottorff JL, Radsma J, Kelly M, Oliffe JL. Fathers' narratives of reducing and quitting smoking. Sociol Health Illn 2009 Mar;31(2):185-200. [doi: 10.1111/j.1467-9566.2008.01126.x] [Medline: 19055591]

8.   Fiore M, Jaen C, Baker T. Treating tobacco use and dependence: 2008 Update. Rockville, MD: U.S. Department of Health and Human Services; 2008.

9.   Oliffe JL, Bottorff JL, Sarbit G. Supporting fathers' efforts to be smoke-free: program principles. Can J Nurs Res 2012 Sep;44(3):64-82. [Medline: 23156192]

10. Robinson M, Robertson S. Young men's health promotion and new information communication technologies: illuminating the issues and research agendas. Health Promot Int 2010 Sep;25(3):363-370 [FREE Full text] [doi: 10.1093/heapro/daq022] [Medline: 20427372]

11. Gwaltney CJ, Metrik J, Kahler CW, Shiffman S. Self-efficacy and smoking cessation: a meta-analysis. Psychol Addict Behav 2009 Mar;23(1):56-66 [FREE Full text] [doi: 10.1037/a0013529] [Medline: 19290690]

12. Lohan M, Aventin A, Maguire L, Clarke M, Linden M, McDaid L. Feasibility trial of a film-based educational intervention for increasing boys' and girls' intentions to avoid teenage pregnancy: Study protocol. Int J Educ Res 2014;68:35-45 [FREE Full text] [doi: 10.1016/j.ijer.2014.08.003] [Medline: 25284956]

13. QuitNow Men. Vancouver, BC: BC Lung Association; 2014. URL: http://men.quitnow.ca/ [accessed 2015-03-26] [WebCite Cache ID 6XKDYblae]

14. Aventin A, Lohan M, O'Halloran P, Henderson M. Design and development of a film-based intervention about teenage men and unintended pregnancy: applying the Medical Research Council framework in practice. Eval Program Plann 2015 Apr;49:19-30 [FREE Full text] [doi: 10.1016/j.evalprogplan.2014.11.003] [Medline: 25483572]

15. Ajzen I. Explaining intentions and behaviour. In: Attitudes, personality and behavior. Maidenhead: Open University Press; 2005.

16. Ajzen I, Madden TJ. Prediction of goal-directed behavior: Attitudes, intentions, and perceived behavioral control. Journal of Experimental Social Psychology 1986 Sep;22(5):453-474. [doi: 10.1016/0022-1031(86)90045-4]

17. Ferguson SG, Frandsen M, Dunbar MS, Shiffman S. Gender and stimulus control of smoking behavior. Nicotine Tob Res 2015 Apr;17(4):431-437. [doi: 10.1093/ntr/ntu195] [Medline: 25762752]

18. Schnoll RA, Martinez E, Tatum KL, Glass M, Bernath A, Ferris D, et al. Increased self-efficacy to quit and perceived control over withdrawal symptoms predict smoking cessation following nicotine dependence treatment. Addict Behav 2011 Feb;36(1-2):144-147 [FREE Full text] [doi: 10.1016/j.addbeh.2010.08.024] [Medline: 20869812]

19. Smit ES, Hoving C, Schelleman-Offermans K, West R, de Vries H. Predictors of successful and unsuccessful quit attempts among smokers motivated to quit. Addict Behav 2014 Sep;39(9):1318-1324. [doi: 10.1016/j.addbeh.2014.04.017] [Medline: 24837754]

Abbreviations

IVD: interactive video drama

Original Paper

Mobile Exercise Apps and Increased Leisure Time Exercise Activity: A Moderated Mediation Analysis of the Role of Self-Efficacy and Barriers

Leib Litman[1], PhD; Zohn Rosen[2], PhD; David Spierer[3], EdD; Sarah Weinberger-Litman[4], PhD; Akiva Goldschein[1]; Jonathan Robinson[5], PhD

[1]Lander College, Psychology Department, Kew Gardens Hills, NY, NY, United States

[2]Mailman School of Public Health, GRAPH Center, Columbia University, New York, NY, United States

[3]Long Island University, Division of Athletic Training, Health and Exercise Science, Brooklyn, NY, United States

[4]Marymount Manhattan College, Psychology Department, New York, NY, United States

[5]Lander College, Computer Science Department, Kew Gardens Hills, NY, NY, United States

Corresponding Author:
Leib Litman, PhD
Lander College
Psychology Department
75-31 150th Street
Kew Gardens Hills, NY, NY, 11367
United States
Phone: 1 917 8367941
Fax: 1 718 820 4800
Email: leib.litman@touro.edu

Abstract

Background: There are currently over 1000 exercise apps for mobile devices on the market. These apps employ a range of features, from tracking exercise activity to providing motivational messages. However, virtually nothing is known about whether exercise apps improve exercise levels and health outcomes and, if so, the mechanisms of these effects.

Objective: Our aim was to examine whether the use of exercise apps is associated with increased levels of exercise and improved health outcomes. We also develop a framework within which to understand how exercise apps may affect health and test multiple models of possible mechanisms of action and boundary conditions of these relationships. Within this framework, app use may increase physical activity by influencing variables such as self-efficacy and may help to overcome exercise barriers, leading to improved health outcomes such as lower body mass index (BMI).

Methods: In this study, 726 participants with one of three backgrounds were surveyed about their use of exercise apps and health: (1) those who never used exercise apps, (2) those who used exercise apps but discontinued use, and (3) those who are currently using exercise apps. Participants were asked about their long-term levels of exercise and about their levels of exercise during the previous week with the International Physical Activity Questionnaire (IPAQ).

Results: Nearly three-quarters of current app users reported being more active compared to under half of non-users and past users. The IPAQ showed that current users had higher total leisure time metabolic equivalent of task (MET) expenditures (1169 METs), including walking and vigorous exercise, compared to those who stopped using their apps (612 METs) or who never used apps (577 METs). Importantly, physical activity levels in domains other than leisure time activity were similar across the groups. The results also showed that current users had lower BMI (25.16) than past users (26.8) and non-users (26.9) and that this association was mediated by exercise levels and self-efficacy. That relationship was also moderated by perceived barriers to exercise. Multiple serial mediation models were tested, which revealed that the association between app use and BMI is mediated by increased self-efficacy and increased exercise.

Conclusions: Exercise app users are more likely to exercise during their leisure time, compared to those who do not use exercise apps, essentially fulfilling the role that many of these apps were designed to accomplish. Data also suggest that one way that exercise apps may increase exercise levels and health outcomes such as BMI is by making it easier for users to overcome barriers to exercise, leading to increased self-efficacy. We discuss ways of improving the effectiveness of apps by incorporating theory-driven

XSL•FO
RenderX

approaches. We conclude that exercise apps can be viewed as intervention delivery systems consisting of features that help users overcome specific barriers.

Introduction

Background

In recent years, the mobile apps market has seen a proliferation of medical and health apps [1-5]—apps whose purpose is to promote health and improve the delivery of health care. Of the wide variety of mobile health apps, exercise and fitness apps are the most popular, accounting for 39% of health-related apps [6].

Lack of exercise puts individuals at an increased risk of numerous chronic health conditions, including metabolic syndrome, cancer, depression, and osteoporosis [7]. It is well documented that sufficient exercise is critical for the promotion of long-term health, weight management, and improved management and prognosis of a variety of chronic illnesses, including a 35% reduction in cardiovascular mortality and a 33% reduction in all-cause mortality [8]. Despite the importance of physical activity, an alarming 30% of people throughout the world are physically inactive [9].

Because of their widespread use, exercise apps have the potential to dramatically improve health outcomes in the United States and around the world, and may thus play an increasingly important role in the public health effort to increase population-wide exercise levels. Empirical evidence is beginning to emerge that the use of exercise-related mobile health technology may be associated with increased exercise levels [10-18] (see [19] for review), and such devices are also beginning to be used in clinical settings [20]. However, the mechanisms by which exercise apps may impact behavior change and health outcomes have not been thoroughly explored.

Current Research

Exercise apps can be viewed as collections of features, each of which has the potential to target specific aspects of cognition, affect, and behavior. Features of exercise apps include providing feedback based on tracking user exercise activity, providing motivational messages, demonstrating the right way to exercise, setting and monitoring goals, incorporating social media, and helping users schedule their exercise program. There are currently over 1000 exercise apps for mobile devices on the market incorporating a myriad of combinations of specific features [6,21]. In this study, we explore how apps may impact behavior at a broad level, by focusing on two variables that are key predictors of physical activity: self-efficacy and barriers to exercise. Most broadly, app features may be conceptualized as targeting specific barriers. To the extent that an app may help an individual overcome specific barriers, it may be expected to increase self-efficacy, leading to increased exercise and, over time, improved health outcomes.

In this study, we test various models of how these variables may be influenced by app use and in turn may help promote increased levels of exercise and health. The models' starting point is the well-established link between exercise and health outcomes. We use body mass index (BMI) as an indicator of health because of its known association with overall health outcomes [22] and its association with exercise levels [23]. We expect to replicate this latter association in this study, in that participants in our sample who report exercising more regularly are expected to have an overall lower BMI. Having replicated that link, we will explore ways that using an exercise app may be associated with changes in BMI, through the effects that app use has on increasing exercise. Specifically, we will test the hypothesis that the use of exercise apps will be associated with greater levels of exercise, which in turn will be associated with lower BMI.

We further predict that individuals who have a high number of barriers to exercise will exercise more when they use an exercise app compared to those who do not use an app. However, individuals who have relatively few barriers will be less likely to benefit from using an exercise-focused app. Figure 1 presents a model of relations between app use, barriers to exercise, exercise levels, and BMI. The model predicts that app use leads to increased exercise (Path 1) and that increased exercise leads to decreased BMI (Path 2). The model also predicts that the effect of app use on exercise (Path 1) will be moderated by barriers. In a separate model, we examine the effect of app use on self-efficacy. App users are expected to have higher self-efficacy, which will mediate exercise levels and in turn mediate BMI.

XSL•FO
RenderX

Figure 1. Model of relations between app use, exercise, and BMI, with barriers to exercise as the moderator.
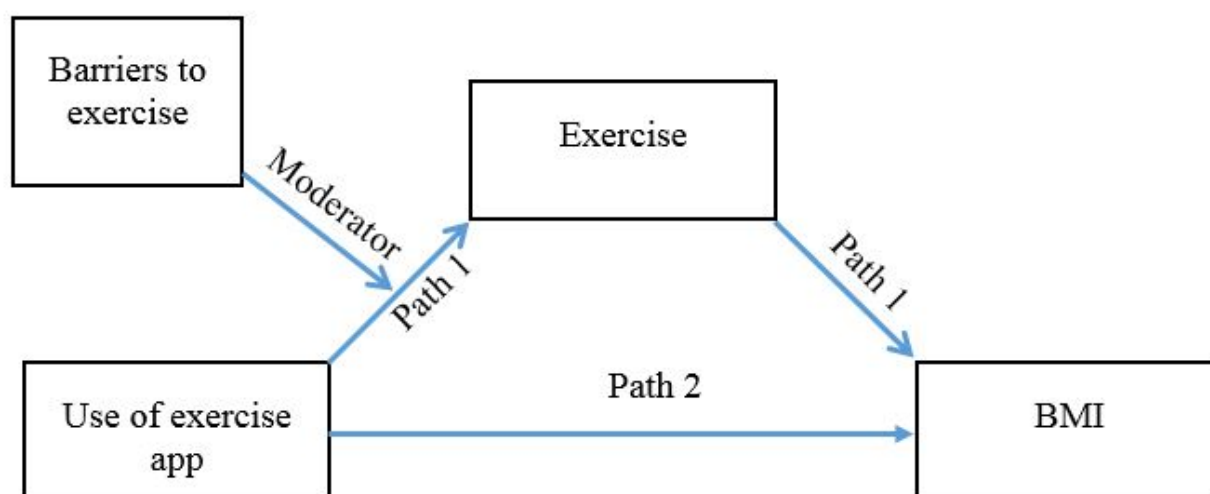
Methods

Participants

A total of 726 participants completed the study; 54.9% (399/726) were male. The dropout rate was 7%. Participants were recruited through Amazon's Mechanical Turk platform. Each participant was paid US $2.50 for completing the study. Participants' ages ranged between 18 and 74 (mean 32.4, SD 11.1).

Design

The study was posted on Amazon Mechanical Turk (MTurk) [24] where a link was made available to a survey, hosted on the Qualtrics website. Participants were asked about their use of exercise apps, and those answers resulted in their assignment to one of three groups: (1) those who never used an exercise app, (2) those who used an exercise app in the past but had since stopped, and (3) those currently using an exercise app. Exercise apps were defined as follows: "Have you ever used an exercise application on any device (iPhone, iPad or Android)? An exercise application is a downloadable App for a mobile device. It is intended to help you exercise (note, apps that help with diet but not exercise are not considered exercise apps)."

The participants who had used an exercise app either currently or in the past were asked questions pertaining to their app use and to the app itself, such as how long they used the app for, if they found it helpful, and what features the app has (eg, social features such as the ability to post results on Facebook, feedback structure, ability to input data). They were also asked to choose their app from a list of apps. In addition, each participant was asked to fill out measures that assessed their level of physical activity, their perceived barriers to and perceived benefits from exercise, their exercise self-efficacy, and self-reported weight and height. The three groups were then compared to see if using an exercise app was associated with our key outcomes, including BMI, frequency of exercise, metabolic equivalent of task (MET) levels, self-efficacy, and barriers to exercise.

The study with all surveys and questions can be found in Multimedia Appendix 1.

Measures

International Physical Activity Questionnaire

Physical activity (PA) was assessed using the International Physical Activity Questionnaire—Long Form (IPAQ) [25]. The IPAQ is a survey that consists of 27 questions on the frequency and duration of PA during the last 7 days. The survey covers five separate domains: (1) job-related physical activity, (2) transportation-related PA such as biking or walking, (3) PA related to housework and gardening, (4) PA done during leisure time, and (5) time spent sitting. Within these activity domains, levels of activity intensity were estimated and assigned an MET: moderate PA=4 MET, vigorous PA=8 MET, walking=3.3 MET, sitting=1 MET. The IPAQ was scored by multiplying the frequency (days per week), duration (minutes per day), and the intensity (MET) of the activities performed, resulting in a MET minutes per week score.

Exercise Benefits/Barriers Scale

The Exercise Benefits/Barriers Scale (EBBS) [26] consists of 43 items with two subscales: one relating to perceived benefits to exercise (29 items) and one relating to perceived barriers of exercise (14 items). The scale has an overall internal reliability of .95, test-retest reliability of .85 [27], and predicts physical activity levels [28].

Exercise Confidence Survey

The Exercise Confidence Survey (ECS) is a 12-item scale that measures exercise self-efficacy [29], using a 5-point Likert scale. The scale has an alpha reliability coefficient of .92, and it significantly predicts self-reported activity [30].

Assessment of Long-Term Exercise Frequency

The IPAQ is limited to assessing activity during a 1-week interval. However, activity levels on any given week can be influenced by multiple factors such as illness, work and family commitments, holidays, and vacations. Therefore we included

a question to assess longer-term exercise patterns that asked participants, "On average, how often have you exercised over the last several months?" The available response options were (1) I don't currently exercise, (2) I am exercising but infrequently (a few times a month at most), (3) I am exercising about once a week, (4) I exercise 2 to 3 times a week, and (5) I exercise more than 3 times a week. Responses 4 and 5 were labeled as "active".

Statistical Analyses

Moderation and Mediation Models

Moderation and mediation models were computed utilizing PROCESS, a mediation and moderation software package [31]. The three app use groups were dummy coded [32], with non-users being the reference group against which current and past users were compared. Barriers to exercise were centered, and cross products of barriers to exercise and app use conditions were computed to produce the two interaction terms required to represent the interaction between the app use condition and barriers to exercise. Exercise frequency was then regressed onto the two contrast variables, with additional covariates included in the model being education level, age, and income.

Mediation

The three levels of app use were dummy coded as in the moderation analysis above. Each of the two dummy-coded variables was entered into the regression equation, with the other dummy-coded variable entered as a covariate, and age as an additional covariate in the model. To test this potential mediation effect, we followed the bootstrapping method, utilizing 5000 iterations [33]. The bootstrapping procedure tests the null hypothesis that the indirect path from the interaction term to the dependent variable via the mediator does not significantly differ from zero. If zero is not contained within the confidence intervals (CI) computed by the bootstrapping procedure, one can conclude that the indirect effect is indeed significantly different from zero at $P<.05$.

Moderated Mediation

Moderated mediation was modeled by entering the contrast between currently using and never used groups (contrast 1) as the predictor, barriers to exercise as the moderator, and BMI as the outcome variable. As described in [31], the contrast between past users and current users groups (contrast 2), the moderator, and the interaction between contrast 1 and the moderator, were entered as covariates. The index of moderated mediation [31] uses the bootstrapping method to test the null hypothesis that the indirect path from the interaction term to the dependent variable via the mediator does not significantly differ from zero.

Descriptive Statistics

Significant age differences in app use were observed. Specifically, participants who were older than 35 years were more likely to have never used an exercise app. Because exercise levels, BMI, and other outcome variables in this study such as attitudes toward exercise and perceived barriers can all differ based on age, we used two separate approaches in controlling for age: (1) exclude participants older than 35 and (2) covarying out the effect of age.

For the purposes of presenting the results descriptively (and the corresponding chi-square tests of independence), we excluded participants older than 35. The purpose of these analyses is to establish the percentage of participants in each app use condition who exercise regularly. Thus, for these analyses, rather than covarying out age, we included only the participants who were younger than 36.

A total barriers score was computed by summing items across all of the barriers examined by the scale. For the purposes of presenting the results descriptively, we dichotomized the barriers scale into high and low barriers individuals, based on the mean of the total barriers score.

For mediation and moderation modeling, all participants, including those above 35, were included in the analyses, and age was entered as a covariate in all models.

Results

Demographic Characteristics of the Sample

Nearly two-thirds of participants (464/726, 63.8%) reported never having used an exercise app (non-users), 15.8% (115/726) reported having used an app in the past but stopping (past users), and 20.2% (147/726) reported using an exercise app currently (current users). Non-users (NU) were older on average (34.3 years old) than past users (PU) (29.1 years old) and current users (CU) (28.1 years old): $F_{2,721}=20.4$, $P<.001$. Sheffe-corrected post-hoc comparisons showed that current and past users did not differ in age from each other ($P>.9$) but both were significantly younger than non-users ($Ps<.001$). There were also differences in BMI between the three groups (NU=26.9, PU=26.7, CU=25.2), $F_{2,721}=4.3$, $P=.014$.

Internal Consistency of Assessment Measures

Table 1 presents zero-order correlations between measures used in this study, which reveal both expected and internally consistent patterns of associations. Participants who reported a lot of barriers to exercise were less active on all measures, had lower self-efficacy, and higher BMI. Self-efficacy was positively correlated with all measures of physical activity, and negatively correlated with BMI. BMI was negatively correlated with length of app use (in weeks) for the current users group ($r_{146}=-.19$, $P<.05$), but not for the past users group ($r_{114}=.05$, $P=.55$).

Exercise frequency (EF) was revealed to be correlated with activity levels as measured by the IPAQ ($r_{724}=.32$ between EF and Total Leisure MET, $P<.001$), thus establishing the convergent validity of the long-term exercise frequency measure. People generally exercise during their leisure times. Thus, exercise levels should correlate strongly with vigorous leisure time activity. Consistent with this approach, exercise frequency was correlated with vigorous leisure MET levels and with total leisure MET levels. However, EF was less strongly correlated with moderate leisure MET levels, and not at all with work MET levels.

Exercise frequency was also correlated with barriers, self-efficacy, and BMI. Although both the exercise frequency and IPAQ measures were significantly correlated with

self-efficacy, barriers, and BMI, exercise frequency was more strongly correlated with these measures than the IPAQ. This is likely due to the IPAQ's exclusive focus on activities during a single week. Because exercise frequency appears to be a more reliable measure of long-term exercise patterns, we use exercise frequency as the outcome variable in our mediation and moderation models. Since the IPAQ can provide insight as to which areas of daily life are likely to be impacted by exercise apps, in subsequent analyses the IPAQ is used to supplement the exercise frequency measure of activity.

Table 1. Zero-order correlations for the total sample (zero-order correlations between all outcome measures, mediators, and moderators).

Measures[a]	1	2	3	4	5	6	7	8
Exercise frequency (1)	–	.32[b]	.34[b]	.22[b]	.01	-.51[b]	.53[b]	-.19[b]
Total leisure MET (2)		–	.93[c]	.65[b]	.14[c]	-.2[b]	.28[b]	.08[c]
Vigorous leisure MET (3)			–	.46[b]	.09	-.22[b]	.28[b]	-.1[c]
Moderate leisure MET (4)				–	.15[c]	-.1[c]	.18[b]	-.05
Work MET (5)					–	-.01	.19[c]	.07
Barriers scale (6)						–	-.53[b]	.27[b]
Self-efficacy scale (7)							–	-.14[b]
BMI (8)								–

[a]Numbers in parentheses correspond to column numbers.

[b]Indicates a significant correlation at the $P<.001$ level.

[c]Indicates a significant correlation at the $P<.05$ level.

Exercise Differences Between App Use Conditions

In the analyses below, we examine whether the use of exercise apps is associated with increased activity levels. As discussed above, for this analysis only participants below age 35 were included. Thus, for the descriptive statistics presented in Tables 2 and 3, all participants were below age 35. These participants did not differ with respect to age: $F_{2,496}=1.6$, $P=.2$.

In the first analysis, we use exercise frequency as the outcome variable. For this analysis, we considered individuals who reported exercising two or more times a week as active. Overall, current users were more likely to be active (73%) compared to non-users (45.8% active), and past users (46.1% active), $\chi^2=30.6$, $P\le.001$. Past users and non-users were not different from each other (see Table 2).

Table 2. Percent of active participants for three app use groups across two levels of barriers.[a]

	Active app users, %					P value		
	Non-users	Past users	Current users	df	X^2	Current users vs non-users	Current users vs past users	Non-users vs past users
All subjects	45.8	46.1	73	2	30.6[b]	<.001	.0001	.53
High barriers	32.2	23.8	60	2	20.7[b]	.001	.001	.44
Low barriers	68.7	72	84.1	2	7.2[c]	.004	.142	.37

[a]Comparisons of current app users, those who began using an app and then stopped, and those who never used an app on self-reported frequency of exercise. Exercise at a rate of twice per week was categorized as active.

[b]Indicates a significant χ^2 at the $P<.001$ level.

[c]Indicates a significant χ^2 at the $P<.05$ level.

We then examined the differences in activity levels across the three app use groups using the IPAQ. Differences in each activity domain were explored using a between-groups analysis of variance (ANOVA). One-tailed planned contrasts were used to compare the current users to non-users and past users, reflecting our a priori hypothesized effect of greater activity levels for the current users. Means, standard deviations, F, and P values for all analyses are presented in Table 3. Overall differences between groups were found for total leisure METs, vigorous leisure METs, and walk leisure METs. Moderate leisure METs were marginally significant. No significant differences in activity levels were found in any other activity domain, including total METs, work METs, transportation to work METs, or home/gardening METs. Planned contrasts further showed that current users had higher activity levels compared to non-users in all leisure subdomains. Current users had higher activity levels compared to past users in walk leisure, and total vigorous leisure METs but not moderate leisure METs. The non-users and past users were not different from each other in any of the leisure activity subdomains.

Table 3. MET comparisons across app use groups.[a]

	Mean (SD)			F [b]	P value	P value		
	Non-users	Past users	Current users			Current user vs non-user	Current user vs past user	Non-user vs past user
Total MET	3724 (4694)	2976 (4382)	4351 (5675)	2.1	.12	.04	.023	.19
Total Leisure MET	577 (1205)	612 (1518)	1169 (2088)	6.4	.002	.001	.009	.85
Vigorous Leisure MET	383 (896)	440 (1005)	730 (1439)	3.8	.02	.006	.034	.67
Walk Leisure MET	191 (395)	146 (521)	427 (729)	8	≤.001	≤.001	.001	.53
Moderate Leisure MET	77 (287)	105 (362)	162 (386)	2.6	.076	.024	.22	.49
Work MET	2976 (3435)	1796 (2378)	2564 (3160)	2.3	.1	.39	.22	.036
Transportation MET	564 (944)	524 (842)	633 (1130)	.277	.76	.56	.48	.77
House Work/ Gardening MET	1450 (2357)	1304 (2282)	1500 (2370)	.19	.83	.55	.85	.61

[a]Comparisons of current app users, those who began using an app and then stopped, and those who never used an app on the IPAQ self-reported metabolic equivalent of task, across eight activity categories. MET values reflect estimated totals for 7 days prior to study participation.

[b]Degrees of freedom are 2 and 512 for all reported omnibus F tests.

Barriers to Exercise

In the next analysis, we examined how the use of exercise apps is associated with barriers to exercise. Percent of active participants, based on the exercise frequency measure in each of the three app groups, along with the omnibus chi-square, P values, and pairwise chi-square comparisons are presented in Table 2 separately for the high and low barriers groups. Low barrier participants were more likely to be active than high barrier participants across all three conditions. Both high and low barrier current app users were significantly more active than non-users. Critically, only the high barrier, but not the low barrier, current users were more active than past users. Past users and non-users were not different from each other independent of their high versus low barrier status. These results suggest that individuals with high barriers to exercise benefit more from using an exercise app, compared to individuals with low barriers to exercise.

One possible explanation for the finding that current app users exercise more than individuals who are not currently using an exercise app is that individuals who exercise more often may be more likely to buy exercise apps. Another possibility is that app use will be associated with more leisure time activity even among active participants who exercise on a regular basis. To examine this, in the next analysis we included only participants who indicated that they are active (ie, exercise at least 2 or 3 times a week). A 2 (high/low barriers) X 3 (non-users, past users and current users) independent-groups ANOVA was then conducted, with Total Leisure MET as the dependent variable. The results revealed a significant main effect of app usage ($F_{2,227}$=3.1, P=.04). Additionally, the app usage X barriers interaction ($F_{2,227}$=3.3, P=.04) remained significant. These results show that individuals with high barriers to exercise benefit more from using exercise apps, even among individuals who exercise regularly.

Simple Moderation: Barriers as Moderators of Exercise

The analyses above suggest that there is an interaction between barriers and app use. To formally model this interaction, we carried out a moderation analysis with app use as a 3-level multi-categorical predictor (non-users, past users, and current users), barriers as the continuous moderator and exercise frequency as the outcome variable.

For all models below, all participants including those older than 35 were included in the analyses, with age as a covariate. The full model accounted for 29.7% of the variance in exercise frequency ($F_{6,700}$=49.2, P<.001). The moderation analysis revealed that the association between app use and exercise levels was significantly moderated by barriers to exercise (R^2 change due to the interaction=.0069, $F_{1,721}$=6.8, P=.009). Specifically, app use was associated with higher levels of exercise for those participants who had more barriers. The Johnson-Neyman bootstrapping technique was used to probe the effect of app use at each level of the continuous moderator (see [34], pp. 234-56). As shown in Figure 2, the association between app use and exercise increases with higher barriers levels. The Johnson-Neyman bootstrapping technique further revealed that a barriers score of 1.45 was the cut-off point below which app use was no longer associated with increased exercise among current users compared to non-users.

The contrast between current users and non-users was the only one that resulted in significant moderation. The contrast between past users and non-users was not significant (R^2 change due to the interaction=.0004, $F_{1,721}$=.3637, P=.55). Most importantly, Johnson-Neyman bootstrapping illustrated that there were no significant changes in the association between app use and exercise levels at any level of the moderator.

Figure 2. Relationship between app use and exercise at different levels of the continuous barriers to exercise moderator.
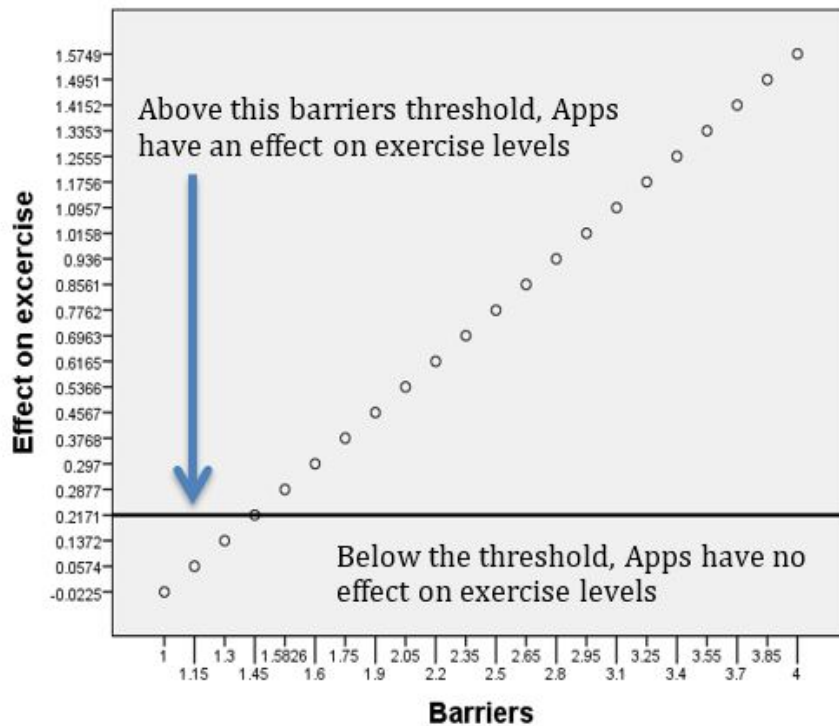
Simple Mediation: Exercise as a Mediator of Body Mass Index

In the next model, we extended the hypothesized effect of app use from exercise to our measure of health status, represented by BMI. In this model (see Figure 3), exercise frequency was entered as a mediator between app use and BMI. The main goal of this mediation model was to test for (1) significant indirect effects of app use on BMI, through exercise levels and (2) to examine whether this indirect effect completely accounts for the association between app use and BMI (ie, no significant direct effect).

The weights for the paths in the model are presented in Figure 3. For the current users versus non-users contrast, the 95% confidence interval for the indirect effect ranged from -1.22 to -.43, indicating that the indirect effect was significantly different from zero at the $P<.05$ level. For the direct effect on the other hand, the 95% confidence interval ranged from -1.47 to 1.15, indicating that the direct effect was not significant. These results show that the reason that app users have lower BMI compared to non-users is that app users exercise more. The lack of a significant direct effect further shows that exercise fully accounts for the BMI differences between current users and those who never used exercise apps.

For the past users versus non-users contrast, the 95% confidence interval for the indirect effect ranged from -.25 to .29, indicating that the indirect effect was not significantly different from zero at the $P<.05$ level. The direct effect was likewise not significant, ranging from -1.14 to 1.48.

Figure 3. Mediation model in which app use was modeled as a 3-level predictor categorical variable in two separate contrasts. Solid lines depict the indirect and direct effects of a 2-level current users vs non-users contrast's effect on BMI. The dashed lines depict the indirect and direct effects effect of a 2-level past users vs non-users contrast's effect on BMI. Asterisk indicates significant results at *P*<.05 level. ns indicates non-significant results at the *P*<.05 level.
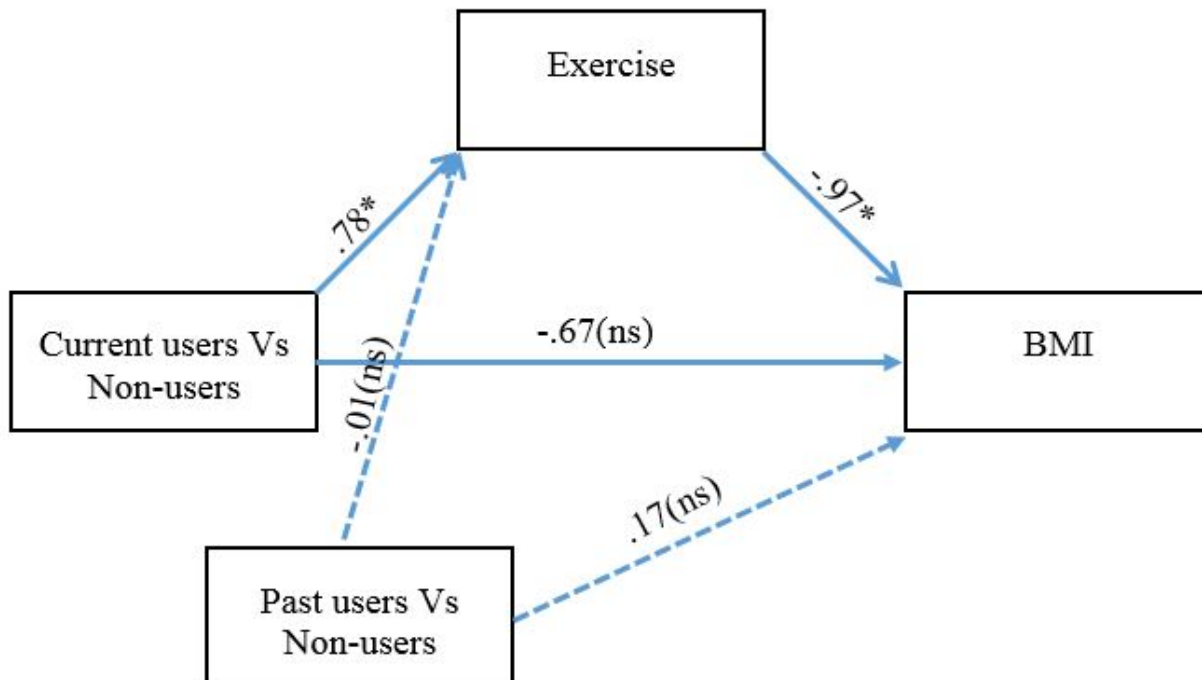
Moderated Mediation: Barriers Moderate the Mediating Effect of Exercise on Body Mass Index

Next, we combined the moderation and mediation models to test a moderated mediation model, in which barriers to exercise moderate the indirect path from app use to BMI through exercise frequency (see Figure 1 above). The moderated mediation model built on the previous analyses showed that (1) app use is associated with decreased BMI, (2) app use is associated with increased exercise frequency, (3) the association between app use and exercise frequency is moderated by barriers to exercise, and (4) that the association of app use and BMI is mediated by exercise levels. The results confirmed the indirect effect of app use on BMI through increased exercise was moderated by barriers to exercise (95% CI -.0821 to -.006).

Modeling Multiple Serial Mediation

The previous analyses established that (1) app use is associated with lower BMI, (2) app use is associated with higher levels of exercise, and (3) that app use is associated with BMI through increased exercise levels. In the next model, we added self-efficacy as an additional plausible mechanism between app use and exercise levels, to examine whether the mediation effect of exercise is itself mediated by self-efficacy. Self-efficacy was added to the model as an additional mediator, which was serial and preceded the exercise frequency mediator in the causal chain (Figure 4). The main goal of this mediation model was to test (1) for serial mediation of the association between app

use and BMI, through self-efficacy and exercise levels, (2) whether self-efficacy mediates the effect directly without exercise in the model, (3) whether this indirect serial mediation effect completely accounts for the association between app use and BMI and (4) whether self-efficacy precedes or follows exercise in the causal chain between app use and BMI.

For the indirect effect of exercise on BMI through self-efficacy and exercise frequency, the 95% confidence interval ranged from -.46 to -.1, indicating that the indirect effect for serial mediation was significantly different from zero at the *P*<.05 level. For the direct effect on the other hand, the 95% confidence interval ranged from -1.87 to .6, indicating that the direct effect was not significant. Reversing the order of the self-efficacy and exercise frequency mediators resulted in non-significant indirect effects for both contrasts.

These results show that app use is associated with increased self-efficacy, which in turn mediates exercise levels, accounting for lower BMI among current app users compared to non-users. Importantly the indirect effect of self-efficacy in the absence of exercise in the model was not significant. This shows that self-efficacy by itself is not a mediator of BMI, but that self-efficacy's mediation of BMI occurs through its mediation of exercise.

When the past users group was entered as the independent variable with the current users groups entered as a covariate, both the direct and indirect effect were not significant (direct

effect: -1.1 to 1.4; indirect effect -0.14 to 0.10). This indicates that past users are not different from non-users.

Figure 4. Mediated mediation model in which app use was modeled as a 3-level predictor categorical variable in two separate contrasts. Solid lines depict the indirect and direct effects of a 2-level current users vs non-users contrast's effect on BMI. The dashed lines depict the indirect and direct effects effect of a 2-level past users vs non-users contrast's effect on BMI. Asterisk indicates significant results at *P*<.05 level. ns indicates non-significant results at *P*<.05 level.
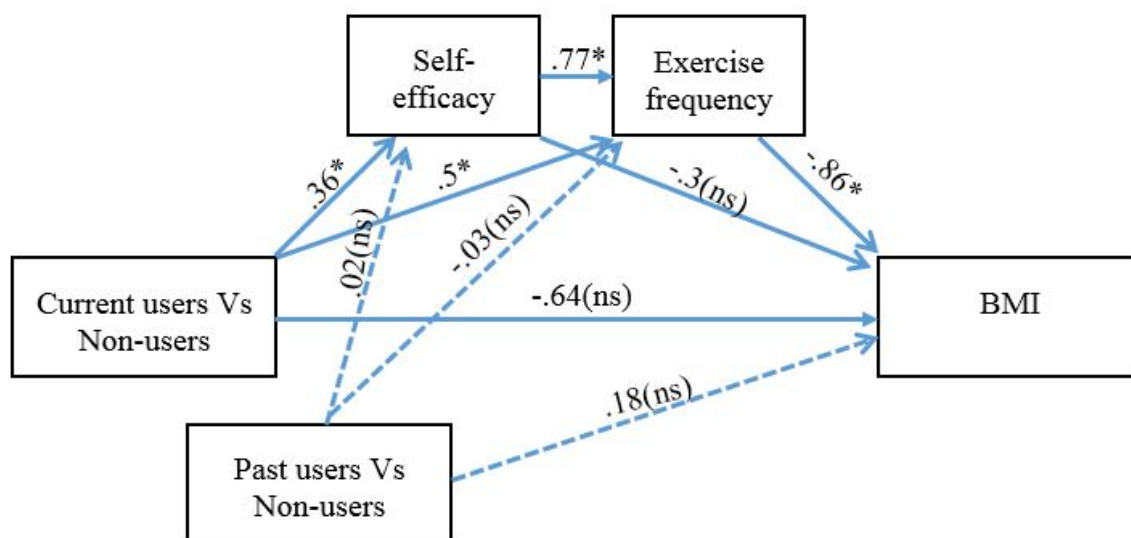
Discussion

Principal Findings

The goal of the current study was to examine whether individuals who use exercise apps exercise more than individuals who do not use such apps, and to identify plausible mechanisms by which the use of exercise apps may lead to increased levels of exercise and improved health outcomes. Current users of exercise apps were 27% more likely to self-report being active compared to participants who have either never used an exercise app or stopped using their apps. Increased activity among app users was specific to leisure time, including vigorous and moderate leisure activities, walking during leisure time, and total leisure time activities. Leisure time activity has been shown in exercise intervention studies to be the most important time for exercise. A meta-analysis of 127 studies showed that the most successful interventions are those that promote increased leisure time activity [35]. Leisure time activity, specifically, also has strong associations with health outcomes including reduced risk of dementia [36], metabolic syndrome [37], myocardial infarction [38], and death [39]. Importantly, no differences were found between current app users and non-users among other activity domains such as total MET or transportation to work. The lack of difference in overall activity shows that app users and non-users in our sample are not different in their overall activity profile, but differ specifically with regard to their leisure time exercise activities. These results show that mobile exercise app users use their leisure time in a more health-oriented way and suggest that mobile exercise apps may help users to increase their leisure time physical exercise activity levels.

Ideally, the three groups in this study represent a sequence of events in time. These results would provide insight as to what occurs when an individual transitions from not being an app user, to being an app user, and then to stopping their app use. Interpreted this way, these results indicate that using an app increases leisure exercise activity and leads to a reduction in BMI, and that stopping to use an app reverses these gains. However, the cross-sectional nature of this study precludes this interpretation. While our analyses utilized covariates such as age, income, and socioeconomic status, it is still possible that pre-existing differences such as interest in exercise can account for the differences between app adopters and non-adopters.

This interpretation is less likely to account for the differences observed between those app users who are currently using their app and those who are no longer using it. In particular, BMI among current users, but not among past users, was negatively correlated with how long they used their exercise app. This result is consistent with the interpretation that extended app use leads the user to exercise more consistently for extended time periods, which leads to reductions in BMI. However, it is still possible that higher interest in exercise, which may be more likely among individuals with lower BMI, can account for both the adoption of an app and the length of its use. It is thus still plausible that individuals who are more interested in exercise are more likely to seek out tools, such as exercise apps, which can help them to achieve their goals—but at the same time these individuals would be more likely to exercise even in the absence of such tools. In other words, app users may be natural "high exercisers", who also have lower BMI.

To control for the potential bias in the disproportionally high number of high exercisers among current app users, we analyzed

a subsample of participants who all reported to be active (ie, exercise at least two times per week). In this subsample of participants, app users still had higher levels of leisure time activity compared to non-users. These analyses reduce the likelihood that pre-existing attitudes toward exercise are able to account for the differences between current app users and non-users that were observed in this study. The models presented in this study put additional significant constraints on the plausibility of this alternative explanation because it was the app users with the highest number of barriers who were more likely to benefit from using an app both in terms of increased exercise levels and decreased BMI. High barrier individuals are generally likely to exercise least, and it is specifically among these individuals that the strongest effects of app use were observed. Although the sum total of our results suggest that there may be a causal relationship between app use and health outcomes, as in any cross-sectional study the possibility of

confounding always exists. Future studies that use experimental and longitudinal study designs will be able to shed more light on the effects described in this study.

Exercise interventions do not directly change behavior [40] but instead have their effect through intermediary variables referred to as mediators, and these effects are often bounded by moderators. To develop an understanding of the mechanisms that lead exercise app users to exercise more, and to explore the possible boundary conditions around these effects, we tested a number of mediation and moderation models. In the mediation models, the indirect effects provide insight as to the plausible mechanisms through which exercise apps, as interventions, may contribute to behavior change. We use the results of these models to develop a framework for understanding the effect of exercise apps as an exercise intervention of potentially high public health value and for contextualizing exercise apps within the framework of behavior change theories (see Figure 5).

Figure 5. Barrier-centered model of exercise apps as exercise behavior-change intervention delivery systems. In this model, an individual's barriers are taken into account during the app design process (b) personality, stages of change, and theory-driven approaches are all used in tailoring sets of features (a) that will maximally help the individual overcome their barriers. The effectiveness of the features leads to increased exercise frequency either directly (d) or through increased self-efficacy (c, e), leading to improved health outcomes (f).
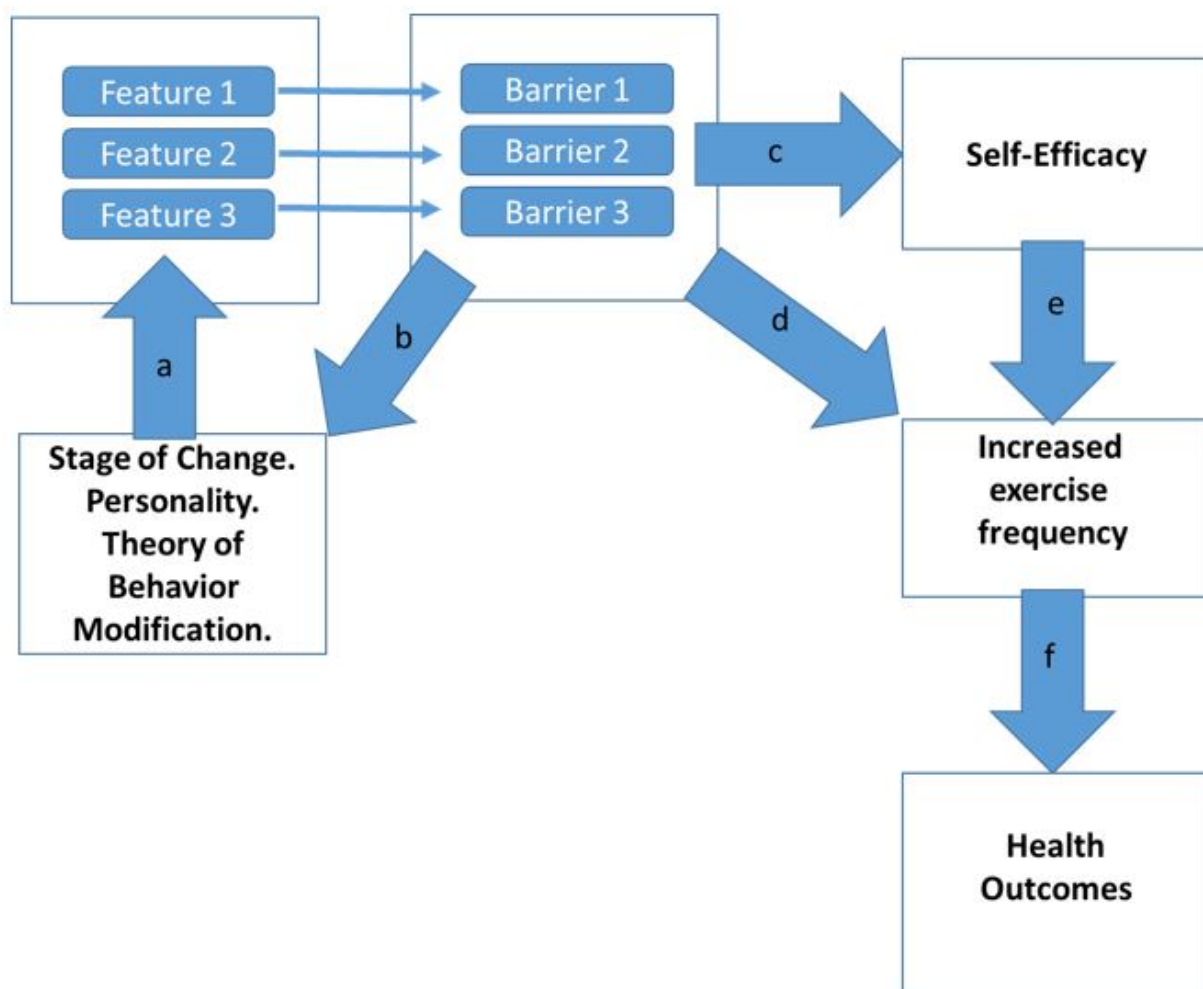
The Role of Barriers

While some people do not exercise due to a lack of interest and motivation, to a large degree lack of exercise is due to specific

barriers that prevent people from exercising [41]. Barriers to exercise can be viewed as the common denominator and the overarching target that all exercise-focused interventions aim to address. For this reason, we chose barriers to exercise as the

starting point in our models. At the broadest level, our models predicted that the effectiveness of apps comes from their ability to help individuals overcome their barriers.

Barriers were found to be significant moderators of app use on exercise. Individuals who have relatively few barriers were found to be less likely to benefit from exercise apps, likely because their engagement in exercise activity is not hampered by barriers, and these individuals are thus less in need of an intervention. Individuals with relatively high levels of barriers on the other hand were most likely to benefit from using exercise apps—an effect that extended to their BMI.

Common barriers to exercise include lack of time, lack of access to exercise facilities, lack of enjoyment of exercise, and/or lack of energy to exercise [42]. For example, an individual may be highly motivated to jog on a regular basis but may be prevented from doing so due to their traveling schedule [43]. Being in a novel location on a regular basis may make it difficult to find appropriate locations for jogging and exercising, thus making participation in consistent physical activity less likely over time. Frequent travel is an objective exercise barrier, and to the extent that an exercise app can provide tools with which to overcome that barrier, it may help the user to exercise more. For example, an app that incorporates a global positioning system (GPS) feature that tracks the user's location and informs the user about nearby places to run or exercise may be able to help that individual overcome that particular barrier. Apps that provide information or video tutorials on proper exercise techniques can help overcome knowledge barriers, and these apps may be especially useful in certain special-needs populations such as pregnant women or individuals with disabilities [44], where special exercise techniques are often required for increased safety. Another example of how apps can help users overcome specific barriers is by features that help to motivate a user to increase their physical activity levels. For example, gamification apps may be able to improve motivation by increasing overall enjoyment of running [45]. Apps can also improve motivation by incorporating motivational messages, especially if those messages are based on tracking individual performance.

Self-Efficacy

Exercise self-efficacy refers to a person's confidence in their ability to engage in exercise [40]. Individuals who are more confident in their ability to exercise have been shown to be more physically active, and numerous studies have shown that belief in one's ability to exercise is one of the best predictors of exercise performance. In a study of over 2500 people, self-efficacy was revealed to be the strongest predictor of vigorous exercise compared to 25 other predictors that were examined [46,47].

Self-efficacy both predicts and is predicted by exercise activity, and is likely to be in a causally reinforcing relationship with exercise [40]. At the cognitive level, interventions that are successful at increasing exercise levels also empower individuals to have confidence in their ability to follow an exercise regimen. This is consistent with two alternative views as to how the use of exercise apps may affect BMI through self-efficacy and exercise. One possibility is that app use increases self-efficacy, which increases exercise levels, leading to lower BMI. Another

possibility is that app use increases exercise levels, which increase self-efficacy. Serial mediation allows for direct testing of these causal paths by changing the order of the serial mediators, because the predictor and outcome variable remain identical in the model and, thus, both models' parameters can be compared directly [48].

Our results show that app use is associated with higher self-efficacy. Self-efficacy mediates the relationship between app use and exercise. Exercise in turn mediates the association between self-efficacy and BMI. Specifically, only the model in which self-efficacy was entered as the first mediator and exercise as the second mediator was significant, but not the model in which the exercise mediator preceded the self-efficacy mediator. Additionally, increased self-efficacy does not directly mediate reductions in BMI, but only through its mediation of increased exercise levels. These results show that self-efficacy is a central mechanism by which the use of exercise apps is associated with both exercise and health outcomes (BMI).

These results provide insight into the possible mechanisms through which apps, as interventions, may influence exercise and health. Self-efficacy interventions have been shown to causally improve exercise levels [49]. Additionally, self-efficacy is effective in helping individuals to overcome their barriers [50]. Increased levels of self-efficacy among app users may stem from the increased confidence the users obtain from using app features that help them overcome their barriers. To the extent that app use helps individuals to overcome their barriers, individuals' self-confidence in their ability to consistently follow an exercise program will increase, and this increase in self-efficacy will be tied to increased exercise levels and positive health outcomes. In summary, our models show that individuals who use apps are more likely to exercise despite their barriers, which mediates increased self-efficacy and reduced BMI.

Exercise Apps and Public Health

Over the last 50 years, the United States population has seen a significant reduction in overall physical activity [51]. Decreased demand on manual labor in the workforce and changes in transportation patterns have contributed to a significant reduction in non-leisure activity levels. While leisure time physical activity has remained largely stable and may even have slightly increased, the net activity levels in the population continue to decrease dramatically [51]. Interventions that can help to at least partially reverse this trend can have an immense impact on overall population-wide health outcomes, and health care expenditures. To that end, the development of effective interventions for increasing overall activity levels are a key aim of public health exercise science and policy.

Mobile apps have a number of advantages over other intervention delivery systems such as Internet websites [52] for promoting physical activity. One advantage of exercise apps is their degree of customizability. An app can incorporate features that are specific for the needs of an individual and, for example, can take individualized data such as connecting to a user's calendar to help schedule exercise sessions. Apps are also mobile and able to provide services on a virtually continuous basis. Because people keep mobile phones with them most of the day, apps can also track activity, monitor progress, and

deliver messages on an ongoing basis. The customizability, versatility, and portability of mobile apps make them ideal as potential intervention delivery systems.

The results presented in this study suggest that apps, as intervention delivery systems, have the potential to significantly improve population exercise levels and may thus have a significant impact on future public health outcomes. Within this framework, apps can be viewed as collections of features. Each feature targets specific barriers, and to the extent that a feature is successful at helping an individual overcome their barriers, it increases their self-efficacy with regard to that barrier and subsequently affects exercise behavior and health outcomes (see Figure 5).

While extant apps provide multiple features that appear to help users overcome their barriers, the effectiveness of exercise apps can be significantly improved by the utilization of individualized, theory-driven approaches (see Figure 5). There is wide recognition that the effectiveness of interventions is enhanced by the theories of behavior change [40]. Reviews of apps that are currently on the market however, reveal a profound lack of theory-driven and evidence-based approaches in app design. For example, in the Transtheoretical Model, barriers differ depending on the stages of change. Individuals in the preparation stage have more perceived barriers compared to those in the active stages. Further, in the active and maintenance stages, barriers that are associated with relapse are going to be

of high importance. However, even the current top exercise apps currently do not have features aimed at relapse prevention [53], although such approaches are currently in development [19]. Additionally, apps should take personality differences into account, as personality traits such as conscientiousness, neuroticism, and extraversion are associated with exercise activity and with specific barriers [54,55]. For example, a low conscientiousness individual who is in the maintenance stage may be more likely to have a scheduling barrier. For this individual, features that address scheduling and incorporate reminders that utilize personalized calendar information, including holidays, appointments, and other commitments have the potential to deliver an individually tailored intervention that takes their needs into account. More research is needed in order to understand how best to incorporate theory into an individualized model of feature delivery. Overall, however, conceptualizing apps as interventions that deliver customizable, theory-driven, person-specific features for the purpose of overcoming barriers provides a helpful framework within which to approach the potential effectiveness of apps in future studies. More work needs to be done to ascertain how to best profile users on a regular basis in order to provide a more individualized feature delivery system and to maximize the match between features and their mode of presentation for each individual. This theory-driven individualized mode of feature presentation is likely to further improve the effectiveness of exercise apps.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Stimulus materials.

[PDF File (Adobe PDF File), 117KB - jmir_v17i8e195_app1.pdf]

References

1. Aungst TD. Medical applications for pharmacists using mobile devices. Ann Pharmacother 2013;47(7-8):1088-1095. [doi: 10.1345/aph.1S035] [Medline: 23821609]

2. Boulos MNK, Wheeler S, Tavares C, Jones R. How smartphones are changing the face of mobile and participatory healthcare: an overview, with example from eCAALYX. Biomed Eng Online 2011;10:24 [FREE Full text] [doi: 10.1186/1475-925X-10-24] [Medline: 21466669]

3. Fox S, Duggan M. Mobile Health. 2012. URL: http://www.pewinternet.org/2012/11/08/mobile-health-2012/ [accessed 2015-07-17] [WebCite Cache ID 6NnAUrVnw]

4. Liu C, Zhu Q, Holroyd KA, Seng EK. Status and trends of mobile-health applications for iOS devices: A developer's perspective. Journal of Systems and Software 2011 Nov;84(11):2022-2033. [doi: 10.1016/j.jss.2011.06.049]

5. Mosa ASM, Yoo I, Sheets L. A systematic review of healthcare applications for smartphones. BMC Med Inform Decis Mak 2012;12:67 [FREE Full text] [doi: 10.1186/1472-6947-12-67] [Medline: 22781312]

6. Aitken M, Gauntlett C. IMS Institute for Healthcare Informatics. 2013. Patient apps for improved healthcare: From novelty to mainstream URL: http://www.imshealth.com/deployedfiles/imshealth/Global/Content/Corporate/IMS%20Health%20Institute/Reports/Patient_Apps/IIHI_Patient_Apps_Report.pdf [accessed 2015-07-17] [WebCite Cache ID 6a5gfebDl]

7. Warburton DER, Nicol CW, Bredin SSD. Health benefits of physical activity: the evidence. CMAJ 2006 Mar 14;174(6):801-809 [FREE Full text] [doi: 10.1503/cmaj.051351] [Medline: 16534088]

8. Nocon M, Hiemann T, Müller-Riemenschneider F, Thalau F, Roll S, Willich SN. Association of physical activity with all-cause and cardiovascular mortality: a systematic review and meta-analysis. Eur J Cardiovasc Prev Rehabil 2008 Jun;15(3):239-246. [doi: 10.1097/HJR.0b013e3282f55e09] [Medline: 18525377]

XSL•FO

RenderX

9.   Hallal PC, Andersen LB, Bull FC, Guthold R, Haskell W, Ekelund U, Lancet Physical Activity Series Working Group. Global physical activity levels: surveillance progress, pitfalls, and prospects. Lancet 2012 Jul 21;380(9838):247-257. [doi: 10.1016/S0140-6736(12)60646-1] [Medline: 22818937]

10.   Fanning J, Mullen S, McAuley E. Increasing physical activity with mobile devices: a meta-analysis. J Med Internet Res 2012;14(6):e161 [FREE Full text] [doi: 10.2196/jmir.2171] [Medline: 23171838]

11.   Fjeldsoe BS, Miller YD, Marshall AL. MobileMums: a randomized controlled trial of an SMS-based physical activity intervention. Ann Behav Med 2010 May;39(2):101-111. [doi: 10.1007/s12160-010-9170-z] [Medline: 20174902]

12.   Kirwan M, Duncan M, Vandelanotte C, Mummery W. Using smartphone technology to monitor physical activity in the 10,000 Steps program: a matched case-control trial. J Med Internet Res 2012;14(2):e55 [FREE Full text] [doi: 10.2196/jmir.1950] [Medline: 22522112]

13.   Lubans DR, Morgan PJ, Dewar D, Collins CE, Plotnikoff RC, Okely AD, et al. The Nutrition and Enjoyable Activity for Teen Girls (NEAT girls) randomized controlled trial for adolescent girls from disadvantaged secondary schools: rationale, study protocol, and baseline results. BMC Public Health 2010;10:652 [FREE Full text] [doi: 10.1186/1471-2458-10-652] [Medline: 21029467]

14.   Newton JR, Marker A, Allen H, Machtmes R, Han H, Johnson WB, et al. Parent-targeted mobile phone intervention to increase physical activity in sedentary children: randomized pilot trial. JMIR Mhealth Uhealth 2014;2(4):e48 [FREE Full text] [doi: 10.2196/mhealth.3420] [Medline: 25386899]

15.   Padmasekara G. Fitness apps, a Valid Alternative to the Gym: a pilot study. JournalMTM 2014 Mar;3(1):37-45. [doi: 10.7309/jmtm.3.1.6]

16.   Prestwich A, Perugini M, Hurling R. Can implementation intentions and text messages promote brisk walking? A randomized trial. Health Psychol 2010 Jan;29(1):40-49. [doi: 10.1037/a0016993] [Medline: 20063934]

17.   Shapiro JR, Bauer S, Hamer RM, Kordy H, Ward D, Bulik CM. Use of text messaging for monitoring sugar-sweetened beverages, physical activity, and screen time in children: a pilot study. J Nutr Educ Behav 2008;40(6):385-391 [FREE Full text] [doi: 10.1016/j.jneb.2007.09.014] [Medline: 18984496]

18.   Sirriyeh R, Lawton R, Ward J. Physical activity and adolescents: an exploratory randomized controlled trial investigating the influence of affective and instrumental text messages. Br J Health Psychol 2010 Nov;15(Pt 4):825-840. [doi: 10.1348/135910710X486889] [Medline: 20156396]

19.   Conroy MB, Yang K, Elci OU, Gabriel KP, Styn MA, Wang J, et al. Physical activity self-monitoring and weight loss: 6-month results of the SMART trial. Med Sci Sports Exerc 2011 Aug;43(8):1568-1574 [FREE Full text] [doi: 10.1249/MSS.0b013e31820b9395] [Medline: 21200337]

20.   Vooijs M, Alpay LL, Snoeck-Stroband JB, Beerthuizen T, Siemonsma PC, Abbink JJ, et al. Validity and usability of low-cost accelerometers for internet-based self-monitoring of physical activity in patients with chronic obstructive pulmonary disease. Interact J Med Res 2014;3(4):e14 [FREE Full text] [doi: 10.2196/ijmr.3056] [Medline: 25347989]

21.   Cowan LT, Van Wagenen SA, Brown BA, Hedin RJ, Seino-Stephan Y, Hall PC, et al. Apps of steel: are exercise apps providing consumers with realistic expectations?: A content analysis of exercise apps for presence of behavior change theory. Health Educ Behav 2013 Apr;40(2):133-139. [doi: 10.1177/1090198112452126] [Medline: 22991048]

22.   Janssen I, Katzmarzyk PT, Ross R. Waist circumference and not body mass index explains obesity-related health risk. Am J Clin Nutr 2004 Mar;79(3):379-384 [FREE Full text] [Medline: 14985210]

23.   Jakicic JM, Clark K, Coleman E, Donnelly JE, Foreyt J, Melanson E, American College of Sports Medicine. American College of Sports Medicine position stand. Appropriate intervention strategies for weight loss and prevention of weight regain for adults. Med Sci Sports Exerc 2001 Dec;33(12):2145-2156. [Medline: 11740312]

24.   Litman L, Robinson J, Rosenzweig C. The relationship between motivation, monetary compensation, and data quality among US- and India-based workers on Mechanical Turk. Behav Res Methods 2015 Jun;47(2):519-528. [doi: 10.3758/s13428-014-0483-x] [Medline: 24907001]

25.   Craig CL, Marshall AL, Sjöström M, Bauman AE, Booth ML, Ainsworth BE, et al. International physical activity questionnaire: 12-country reliability and validity. Med Sci Sports Exerc 2003 Aug;35(8):1381-1395. [doi: 10.1249/01.MSS.0000078924.61453.FB] [Medline: 12900694]

26.   Sechrist KR, Walker SN, Pender NJ. Development and psychometric evaluation of the exercise benefits/barriers scale. Res Nurs Health 1987 Dec;10(6):357-365. [Medline: 3423307]

27.   Steinhardt MA, Dishman RK. Reliability and validity of expected outcomes and barriers for habitual physical activity. J Occup Med 1989 Jun;31(6):536-546. [Medline: 2786559]

28.   Brown SA. Measuring perceived benefits and perceived barriers for physical activity. Am J Health Behav 2005;29(2):107-116. [Medline: 15698978]

29.   Sallis JS, Pinski RB, Grossman RM, Patterson TL, Nader PR. The development of self-efficacy scales for health-related diet and exercise behaviors. Health Educ Res 1988;3(3):283-292. [doi: 10.1093/her/3.3.283]

30.   Resnick B, Jenkins LS. Testing the reliability and validity of the Self-Efficacy for Exercise scale. Nurs Res 2000;49(3):154-159. [Medline: 10882320]

31.   Hayes A. White paper. A versatile computational tool for observed variable mediation, moderation, and conditional process modeling URL: http://www.afhayes.com/public/process.pdf [accessed 2015-07-17] [WebCite Cache ID 6a5hwRhAT]

32. Hayes AF, Preacher KJ. Statistical mediation analysis with a multicategorical independent variable. Br J Math Stat Psychol 2014 Nov;67(3):451-470. [doi: 10.1111/bmsp.12028] [Medline: 24188158]

33. Preacher KJ, Rucker DD, Hayes AP. Addressing moderated mediation hypotheses: Theory, methods, and prescriptions. Multivariate behavioral research 2007;42(1):185-227.

34. Preacher KJ, Hayes AF. SPSS and SAS procedures for estimating indirect effects in simple mediation models. Behav Res Methods Instrum Comput 2004 Nov;36(4):717-731. [Medline: 15641418]

35. Dishman RK, Buckworth J. Increasing physical activity: a quantitative synthesis. Med Sci Sports Exerc 1996 Jun;28(6):706-719. [Medline: 8784759]

36. Rovio S, Kåreholt I, Helkala E, Viitanen M, Winblad B, Tuomilehto J, et al. Leisure-time physical activity at midlife and the risk of dementia and Alzheimer's disease. Lancet Neurol 2005 Nov;4(11):705-711. [doi: 10.1016/S1474-4422(05)70198-8] [Medline: 16239176]

37. Laaksonen DE, Lakka HM, Salonen JT, Niskanen LK, Rauramaa R, Lakka TA. Low levels of leisure-time physical activity and cardiorespiratory fitness predict development of the metabolic syndrome. Diabetes Care 2002 Sep;25(9):1612-1618. [Medline: 12196436]

38. Lakka TA, Venäläinen JM, Rauramaa R, Salonen R, Tuomilehto J, Salonen JT. Relation of leisure-time physical activity and cardiorespiratory fitness to the risk of acute myocardial infarction. N Engl J Med 1994 Jun 2;330(22):1549-1554. [doi: 10.1056/NEJM199406023302201] [Medline: 8177243]

39. Leon AS, Connett J, Jacobs DR, Rauramaa R. Leisure-time physical activity levels and risk of coronary heart disease and death. The Multiple Risk Factor Intervention Trial. JAMA 1987 Nov 6;258(17):2388-2395. [Medline: 3669210]

40. Sallis J, Owen N. Physical activity & behavioral medicine. Thousand Oaks, Calif: Sage Publications; 1999.

41. Huppertz C, Bartels M, Jansen IE, Boomsma DI, Willemsen G, de Moor HM, et al. A twin-sibling study on the relationship between exercise attitudes and exercise behavior. Behav Genet 2014 Jan;44(1):45-55 [FREE Full text] [doi: 10.1007/s10519-013-9617-7] [Medline: 24072598]

42. Ebben W, Brudzynski L. Motivations and barriers to exercise among college students. Journal of Exercise Physiology Online 2008 Oct;11(5):1-11.

43. Estabrooks P, Carron AV. The conceptualization and effect of control beliefs on exercise attendance in the elderly. J Aging Health 1998 Nov;10(4):441-457. [Medline: 10346694]

44. Rimmer JH, Riley B, Wang E, Rauworth A, Jurkowski J. Physical activity participation among persons with disabilities: barriers and facilitators. Am J Prev Med 2004 Jun;26(5):419-425. [doi: 10.1016/j.amepre.2004.02.002] [Medline: 15165658]

45. Yim J, Graham T. Using games to increase exercise motivation. : ACM; 2007 Presented at: Conference on Future Play; November 15-17, 2007; Toronto, Canada p. 166-173.

46. Sallis JF, Hovell MF, Hofstetter CR. Predictors of adoption and maintenance of vigorous physical activity in men and women. Prev Med 1992 Mar;21(2):237-251. [Medline: 1579558]

47. Sallis JF, Hovell MF, Hofstetter CR, Faucher P, Elder JP, Blanchard J, et al. A multivariate study of determinants of vigorous exercise in a community sample. Prev Med 1989 Jan;18(1):20-34. [Medline: 2710760]

48. Hayes A. Introduction to mediation, moderation, and conditional process analysis: A regression-based approach. New York: The Guilford Press; 2013.

49. McAuley E, Blissmer B. Self-efficacy determinants and consequences of physical activity. Exerc Sport Sci Rev 2000 Apr;28(2):85-88. [Medline: 10902091]

50. Lee L, Arthur A, Avis M. Using self-efficacy theory to develop interventions that help older people overcome psychological barriers to physical activity: a discussion paper. Int J Nurs Stud 2008 Nov;45(11):1690-1699. [doi: 10.1016/j.ijnurstu.2008.02.012] [Medline: 18501359]

51. Brownson RC, Boehmer TK, Luke DA. Declining rates of physical activity in the United States: what are the contributors? Annu Rev Public Health 2005;26:421-443. [doi: 10.1146/annurev.publhealth.26.021304.144437] [Medline: 15760296]

52. Doshi A, Patrick K, Sallis JF, Calfas K. Evaluation of physical activity web sites for use of behavior change theories. Ann Behav Med 2003;25(2):105-111. [Medline: 12704012]

53. Direito A, Dale LP, Shields E, Dobson R, Whittaker R, Maddison R. Do physical activity and dietary smartphone applications incorporate evidence-based behaviour change techniques? BMC Public Health 2014;14:646 [FREE Full text] [doi: 10.1186/1471-2458-14-646] [Medline: 24965805]

54. Courneya K, Hellsten L. Personality correlates of exercise behavior, motives, barriers and preferences: An application of the five-factor model. Personality and Individual Differences 1998 May;24(5):625-633. [doi: 10.1016/S0191-8869(97)00231-6]

55. O'Connor DB, Conner M, Jones F, McMillan B, Ferguson E. Exploring the benefits of conscientiousness: an investigation of the role of daily stressors and health behaviors. Ann Behav Med 2009 Apr;37(2):184-196. [doi: 10.1007/s12160-009-9087-6] [Medline: 19322619]

Abbreviations

ANOVA: analysis of variance

XSL•FO

RenderX

BMI: body mass index
CI: confidence intervals
CU: current users
EBBS: Exercise Benefits/Barriers Scale
ECS: The Exercise Confidence Survey
EF: exercise frequency
GPS: global positioning system
IPAQ: International Physical Activity Questionnaire
MET: metabolic equivalent of task
MTurk: Amazon Mechanical Turk
NU: non-users
PA: physical activity
PU: past users

An Interactive Computer Session to Initiate Physical Activity in Sedentary Cardiac Patients: Randomized Controlled Trial

Fabio A Almeida[1], MSW, PhD; Renae L Smith-Ray[2], PhD; David A Dzewaltowski[3], PhD; Russell E Glasgow[4], PhD; Rebecca E Lee[5], PhD; Deborah SK Thomas[6], PhD; Stanley Xu[7], PhD; Paul A Estabrooks[8], PhD

[1]Implementation and Systems Science Laboratory, Department of Human, Nutrition, Foods, and Exercise, Virginia Tech, Roanoke, VA, United States

[2]Center for Research on Health and Aging, Institute for Health Research and Policy, University of Illinois at Chicago, Chicago, IL, United States

[3]Department of Kinesiology, Kansas State University, Manhattan, KS, United States

[4]Department of Family Medicine, University of Colorado School of Medicine, Aurora, CO, United States

[5]College of Nursing and Health Innovation, Arizona State University, Tempe, AZ, United States

[6]Department of Geography and Environmental Sciences, University of Colorado Denver, Denver, CO, United States

[7]Institute for Health Research, Kaiser Permanente Colorado, Aurora, CO, United States

[8]Department of Human Nutrition, Foods and Exercise, Virginia Tech and Carilion Clinic, Roanoke, VA, United States

Corresponding Author:
Fabio A Almeida, MSW, PhD
Implementation and Systems Science Laboratory
Department of Human, Nutrition, Foods, and Exercise
Virginia Tech
Suite #104
1 Riverside Circle SW
Roanoke, VA, 24073
United States
Phone: 1 540 231 6198
Fax: 1 540 857 6658
Email: falmeida@vt.edu

Abstract

Background: Physical activity (PA) improves many facets of health. Despite this, the majority of American adults are insufficiently active. Adults who visit a physician complaining of chest pain and related cardiovascular symptoms are often referred for further testing. However, when this testing does not reveal an underlying disease or pathology, patients typically receive no additional standard care services. A PA intervention delivered within the clinic setting may be an effective strategy for improving the health of this population at a time when they may be motivated to take preventive action.

Objective: Our aim was to determine the effectiveness of a tailored, computer-based, interactive personal action planning session to initiate PA among a group of sedentary cardiac patients following exercise treadmill testing (ETT).

Methods: This study was part of a larger 2x2 randomized controlled trial to determine the impact of environmental and social-cognitive intervention approaches on the initiation and maintenance of weekly PA for patients post ETT. Participants who were referred to an ETT center but had a negative-test (ie, stress tests results indicated no apparent cardiac issues) were randomized to one of four treatment arms: (1) increased environmental accessibility to PA resources via the provision of a free voucher to a fitness facility in close proximity to their home or workplace (ENV), (2) a tailored social cognitive intervention (SC) using a "5 As"-based (ask, advise, assess, assist, and arrange) personal action planning tool, (3) combined intervention of both ENV and SC approaches (COMBO), or (4) a matched contact nutrition control (CON). Each intervention was delivered using a computer-based interactive session. A general linear model for repeated measures was conducted with change in PA behavior from baseline to 1-month post interactive computer session as the primary outcome.

Results: Sedentary participants (n=452; 34.7% participation rate) without a gym membership (mean age 58.57 years; 59% female, 78% white, 12% black, 11% Hispanic) completed a baseline assessment and an interactive computer session. PA increased across the study sample ($F_{1,441}=30.03$, $P<.001$). However, a time by condition interaction ($F_{3,441}=8.33$, $P<.001$) followed by post hoc analyses indicated that SC participants exhibited a significant increase in weekly PA participation (mean 45.1, SD 10.2) compared to CON (mean -2.5, SD 10.8, $P=.004$) and ENV (mean 8.3, SD 8.1, $P<.05$). Additionally, COMBO participants exhibited

a significant increase in weekly PA participation (mean 53.4, SD 8.9) compared to CON ($P<.001$) and ENV ($P=.003$) participants. There were no significant differences between ENV and CON or between SC and COMBO.

Conclusions: A brief, computer-based, interactive personal action planning session may be an effective tool to initiate PA within a health care setting, in particular as part of the ETT system.

Trial Registration: Clinicaltrials.gov NCT00432133, http://clinicaltrials.gov/ct2/show/NCT00432133 (Archived by WebCite at http://www.webcitation.org/6aa8X3mw1).

KEYWORDS

exercise, physical; treadmill test; human computer interaction; behavioral research; cardiovascular diseases; interactive media

Introduction

Physical inactivity is a known contributing factor to many negative health conditions including cardiovascular disease, obesity, cancer, and diabetes [1,2]. Regular physical activity (PA) leads to improved exercise capacity, improved heart rate, and reduces the risk of cardiovascular disease and mortality [3,4]. Despite this, less than one third of adults in the United States engage in a sufficient amount of leisure time PA according to the Healthy People 2020 guidelines [5]. Regular PA is recommended by the American College of Sport Medicine and the American Heart Association (AHA) as tertiary prevention for older adults who have been diagnosed with coronary heart disease (CHD) and hypertension [4,6].

As a result, both individual-level and environmental-level strategies have been used to promote PA behavior change [7-10]. Many of these strategies have used social cognitive theory as a theoretical basis to help predict and improve individual-level PA participation [7,11]. On the other hand, social ecological theories of PA claim that individual-level factors alone are not sufficient predictors of PA participation; instead, environmental accessibility also plays an important role [9,10]. While it is clear that different theoretical underpinnings help predict PA participation, it is unclear whether individual or environmental factors have a greater impact on short-term PA outcomes.

In addition to having a strong theoretical basis, the intervention setting also plays an important part on PA behavior change. A review of 73 studies commissioned by the US Preventive Task Force (USPTF) concluded that behavioral counseling by physicians and other health facility-based practitioners that includes PA is associated with improving indicators of cardiac health [12]. However, studies show that both patients and physicians reported that less than 50% of patients ever receive any PA counseling as part of their regular office visits [13-16]. This could be due to competing demands, lack of PA training and or knowledge by physicians, or time constraints within a busy health care practice. For these reasons, the feasibility of incorporating a PA intervention into the health care delivery system is challenging. Furthermore, a Cochrane review showed that tailored rather than general PA information is more likely to help patients engage in PA, but that tailored PA information coupled with information on outside exercise resources was more effective than brief physician counseling on PA alone [17].

Limited information is available on feasible approaches for delivering tailored, clinic-based PA counseling that includes resources for PA outside of the health care system. In Australia, Eakin et al have demonstrated that PA can be increased when health care settings use telephone-counseling approaches [18]. In the United States, the Activity Counseling Trial [19] used a combination of physician advice and onsite health educators with the telephone support to successfully improve cardiorespiratory fitness among women with no improvement reported among men. However, these approaches have not been translated to clinical settings in the United States and are still resource intensive for delivery across a population of patients.

On the other hand, technology-based approaches have the potential to enhance feasibility and dissemination of this type of clinic-based PA intervention [20]. Technology-based interventions allow for rapid dissemination of health messages using one or more forms such as computer/Internet, smart phone apps, text messaging, and interactive voice response [21,22]. Technology-based interventions are a resource-efficient and affordable delivery method that enables individual tailoring and adaptation, frequent assessment, automated treatment delivery, and the potential to reach large samples [21-24]. Interactive computer sessions have been used to target nutrition and PA knowledge, dietary patterns, PA participation, and exercise self-efficacy in various settings [25-29].

Early evidence for the effectiveness of tailored technology-based PA programs was mixed [30-32], but more recent research suggests that eHealth interventions have a significant positive impact on PA participation and quality of life, and they improve maintenance of health behavior change [22,29,33-35]. Many eHealth studies have faced methodological challenges including accessibility of mobile devices such as smartphones, high attrition, and self-reported measurement [22,30,31,34]. However, Mounton and Cloes found that, compared to a Web-based intervention, a center-based PA intervention was superior for improving PA participation, but the Web-based intervention was more beneficial for increasing health education [36]. Moreover, Compernolle et al found that a tailored, Web-based PA intervention was most effective for participants who were insufficiently active at baseline [33]. Together, this evidence suggests that eHealth interventions have the potential to lead to effective health behavior change using tailored messages delivered through resource-efficient platforms. Although the USPTF review found improvements in some measures of cardiac health following clinician-based PA counseling [12], there continues to be a need for additional development and testing

of feasible low-dosage PA interventions within health care delivery systems.

Exercise treadmill testing (ETT) is often one of the first diagnostic procedures performed on patients who present to their physicians with chest pain [37]. For those patients with positive test results, standard practice care follows, which includes consultations with cardiac specialists, intensive risk factor management, and careful clinical follow-up. For those patients with negative test results, there are no additional standard care services provided and often they are sent home with advice to monitor their dietary intake and exercise more. These patients may develop a perception of low risk of future cardiac events; however, these cardiac patients remain at a higher risk of mortality when compared to the general population [38]. A PA intervention delivered within the clinical setting may be an effective strategy for improving the health of this population at a time when they may be motivated to take preventive action.

The current study fills these gaps by testing the impact of a clinic-based, tailored, interactive computer session on PA participation using individual and environmental strategies. Most of the previous studies examined were not developed and integrated as part of the health care delivery system [12,30-32,39], particularly the ETT system, which presents an optimal opportunity for PA counseling for a high-risk population. This study used an integrated research-practice systems-based approach [40] and sought to incorporate the USPTF recommendation following negative ETT by developing and testing a low- to medium-intensity, theory-based, tailored PA intervention, delivered within the health care system.

The purpose of this study was to determine the effectiveness of a tailored, computer-based interactive personal action planning session to initiate PA among a group of sedentary patients. It was hypothesized that compared to controls, patients who completed the computer-based interactive personal action planning session, or received a voucher for a local fitness facility, or received both would significantly increase PA participation.

Methods

Design

This paper presents physical activity initiation information as part of a larger trial—CardiACTION, a 2x2 randomized controlled trial to test the efficacy of three PA interventions compared to a matched contact control group [41]. Participants were randomly assigned to one of four treatment arms: (1) increased environmental accessibility to PA resources (ENV), (2) a tailored, social cognitive PA intervention operationalized using the "5 As" (ask, advise, assess, assist, and arrange) (SC) [18,31,34], (3) combined PA intervention of both ENV and SC approaches (COMBO), or (4) a matched contact nutrition control (CON). For the larger trial, we powered our study to detect statistically significant changes in minutes of moderate to vigorous intensity PA at 6 months between conditions. For a more conservative estimate of the appropriate sample size, a Bonferroni correction was used by dividing alpha by the number

of comparisons necessary with the 2x2 design. As such, to detect a 90-minute difference between conditions with a power of 95% and an alpha of .0083, a total of 93 participants per group were needed.

Participants

Adults age 18 and over were recruited to the CardiACTION trial between April 2004 and April 2006 during an outpatient health care visit in which they completed an ETT. All patients were referred to a cardiac stress test by their primary care physician for diagnosis following symptoms related to cardiovascular disease (eg, heart palpitations, shortness of breath, chest pain) in an integrated health care system in the United States. The inclusion criteria for this study were (1) no chest pains and a normal electrocardiogram during the treadmill stress test and (2) currently physically inactive or insufficiently active (<150 moderate intensity PA/week). Patients were excluded from the study if they (1) were younger than 18 years of age, (2) had no access to a telephone, (3) were not able to read or understand English, (4) had a contraindication to PA identified during stress testing, and (5) already had a membership to a recreation center.

A total of 4097 patients completed ETT during the study recruitment period and 1300 participants met eligibility criteria (31.73%). The majority of patients were excluded because they reported sufficient PA (1507/4097, 36.78%), had a positive ETT (631/4097, 15.40%), had a gym membership (243/4097, 5.93%), were not able to read or understand English (138/4097, 3.37%), or had a contraindication to PA identified during stress testing (121/4097, 2.95%). Finally, a total of 157 patients were not approached for participation. Of those eligible who declined to participate (n=848), the most common reasons cited for declining included not being interested (n=314), being too busy (n=201), or not wanting to participate in research (n=119). A total of 452 patients (34.7% of eligible patients) completed a written informed consent, baseline assessments, and were enrolled in the study. Those eligible who enrolled in the study were more likely to be female (59.9% vs 47.0%) and African American (12.6% vs. 7.1%) when compared to those eligible who declined participation. There were no differences based on ethnicity, age, or PA between those who agreed and those who declined participation. Those enrolled were then randomly assigned to one of the four interventions: ENV, SC, COMBO, or CON. Group assignment was stratified based on gender (40% female) and abnormal heart rate recovery results (33% abnormal) following a randomization table created by the study statistician. Trained research assistants opened a pre-determined envelope with group assignment upon completion of all baseline measurements. To reduce potential bias, research assistants were blinded to the contents of the envelope.

Procedures

All patients who met preliminary eligibility criteria were informed of the CardiACTION study by an exercise physiologist at the completion of the ETT. Those who showed interest in the study were referred to an onsite research assistant who explained the study in detail and conducted a short screening to determine eligibility. Patients eligible to participate in the study completed the informed consent process and a computer-based survey. A subsequent study visit at the treadmill center was then scheduled

for 10 days after initial recruitment to complete the randomization process and initiate the study intervention. During the 10-day follow-up visit, all participants were randomized to one of the four interventions after completion of all baseline assessments and subsequently completed an interactive computer session specific to their intervention arm. All participants, regardless of condition, were matched on contact frequency and type. Specifically, following the computer interactive session, all participants received 3 interactive voice response (IVR) support calls and 3 tailored newsletters at 1, 3, and 5 months post randomization (see [41] for details). ENV group participants' computer interactive session included an interactive geographic information system (GIS) interface that allowed participants to select a free 12-month voucher to a fitness facility in close proximity to their home or workplace or the path in between home and work. SC group participants' session included personal action planning to improve self- and response-efficacy as well as personal goals for PA. COMBO group participants' session included both the ENV and SC components. CON group participants' session targeted goals and personal action planning strategies for healthy eating.

The interactive computer session, its components, development and theoretical underpinnings, and all subsequent intervention components have been described elsewhere [41]. Briefly, the interactive computer session allowed the use of audio information, visual graphics, text, video, and the use of familiar models and locations. Videos for all four arms took approximately 20 minutes to complete and were filmed in the clinic's treadmill center and featured the clinic's Chief of Cardiology. The SC, ENV, and COMBO interventions began with an opening message about the importance of PA to achieving good health, while the CON intervention focused on the importance of healthy eating.

Following the opening message, participants in the SC and COMBO group proceeded through an interactive action planning session based on the 5 As [42]. The session included an *assessment* of the patient's PA level and motives for increasing PA and provided *advice* on the recommended level of PA. The interactive computer program then provided a range of minutes of PA that would be an appropriate starting point for the patient and used a collaborative goal setting process to *agree* on a goal for the upcoming month. Finally, the interactive session included *assisting* the patients with PA barrier identification and strategies to overcome barriers, followed by *arranging* for follow-up with an automated telephone counseling call 1 month later [42]. The CON group followed the same process with a focus on healthy eating (ie, personal action plan to increase fruit and vegetable consumption and decrease fat intake). The ENV group participants proceeded through a mapping tool (GIS) designed to help them identify fitness facilities close to their homes and/or work. Participants were able to look at facilities by location and amenities available. At the end of the session, participants in the ENV group printed their free 12-month membership voucher to their facility of choice. The COMBO group completed the GIS mapping tool after completing a personal action plan using the 5 As, as described above.

One month after the completion of the interactive computer session, all participants received a follow-up interactive voice response (IVR) automated telephone call. The IVR system pulled individual information (ie, PA and eating behavior, goals, strategies, and barriers selected) directly from the CD-ROM session and called participants at designated times on multiple days over a week to increase the likelihood of participant completion. All participants completed a PA assessment as part of the IVR follow-up. The study design and protocol was approved by the Kaiser Permanente Colorado Institutional Review Board and is registered (NCT00432133).

Measures

Physical activity was defined as the total minutes of moderate and vigorous PA performed each week, measured using PA participation questions from the Centers for Disease Control and Prevention Behavioral Risk Factor Surveillance Survey [43]. This measurement tool has shown validity and reliability in determining respondent PA levels [44,45]. This tool requires participants to respond to a series of questions regarding PA intensity (light, moderate, and vigorous), frequency (number of days per week), and duration (number of minutes per day) in a typical week. The PA measure was integrated into the computer interactive session as well as the IVR follow-up call. These responses were used to calculate total weekly minutes of moderate and vigorous PA and the percentage of study participants meeting current recommendations for 150 minutes or more of moderate and vigorous PA per week.

Statistical Analysis

Statistical analysis (IBM SPSS v21.0) included descriptive statistics of sex, age, race, ethnicity, and baseline minutes of moderate and vigorous physical activity. Chi-square and independent *t* tests were conducted to determine if any of the groups differed on baseline characteristics (Table 1). A general linear model for repeated measures analysis was conducted using a within-subjects factor of time (baseline and 1-month follow-up), between-subjects factor of group (ENV, SC, COMBO, and CON), and change in weekly PA minutes between baseline and 1-month as the primary dependent variable. Post hoc analyses were conducted to investigate any significant effects found with alpha set at .05.

Additionally, the percentage of all participants currently meeting public health recommendations for aerobic exercise at 1-month follow-up was calculated by group. A dichotomous variable was developed to indicate whether participants were meeting public health recommendations of 150 minutes moderate to vigorous PA minutes per week. Logistic regression models were conducted to determine if study group had a significant impact on attainment of PA public health recommendations.

Finally, we used intention-to-treat (ITT) analysis to include all participants with non-missing baseline outcome measurements (7 participants had missing baseline values). For those participants with missing 1-month follow-up PA measurement (37/452, 8.18%), we replaced the missing 1-month data with their baseline value following the Last Observation Carried Forward approach. We also conducted subgroup analysis including only those participants who cashed their vouchers in the COMBO (67/125, 53.6%) and ENV (63/115, 54.78%) groups, and our results remained unchanged. As such, we

present the results of our final ITT analysis with 445 participants.

Results

Participant Characteristics

Figure 1 depicts the flow diagram of participants through the different stages of the study. In total, 452 cardiac patients were enrolled in the study and completed baseline assessments and the interactive computer session. The average age of the participants was 58.6 years (SD 9.6) and 59.9% were women. Over three-quarters of the sample was white and 12% and 10% of the sample were African American and Latino, respectively. At 1-month follow-up, 91.8% (415/452) of participants completed the automated phone calls and provided data included

in the analysis. Although the remaining 8.2% (37/452) of participants did not complete the 1-month follow-up IVR call, no participants had actively withdrawn from the study. Participants who were randomized to the SC group ($F_{1,3}$=5.254, P<.001; SC=83%, ENV=96.5%, COMBO=92.8%, CON=94.3%) were significantly less likely to complete the 1-month IVR follow-up than their counterparts. No significant differences in non-completion were found among the other groups. Furthermore, there were no significant differences in 1-month IVR call completion by sex, education, race, ethnicity, marital status, income, employment, age, and weekly PA participation, both within and between groups. Baseline characteristics for all study participants are shown on Table 1. Baseline measurements did not differ significantly between the four intervention groups.

Table 1. Participant characteristics at baseline.[a]

Participant characteristics	Environmental (n=115)	Social cognitive (n=106)	Combination (n=125)	Control (n=106)	Overall (n=452)
Age, mean (SD)	58.77 (9.49)	59.62 (9.48)	58.06 (10.12)	57.92 (9.55)	58.57 (9.67)
Female, %	53.9	61.3	60	62.3	59.3
Black, %	13.2	9.5	12.9	11.3	11.8
Hispanic or Latino, %	9.6	11.4	12.1	12.3	11.4
High school or less, %	36	39	44.4	40.6	40.1
Married, %	69.3	60	69.4	67	66.6
USD $30,000 or less, %	10.5	14.3	8.9	15.1	12
Employed full-time, %	66.7	67.6	67.7	64.2	66.6
Moderate and vigorous minutes of PA, mean (SD)	65.77 (65.3)	63.38 (78.2)	56.25 (66.72)	80.50 (89.1)	65.9 (75.1)

[a]No significant differences between groups.

Figure 1. Flow diagram of participants' progress through study.
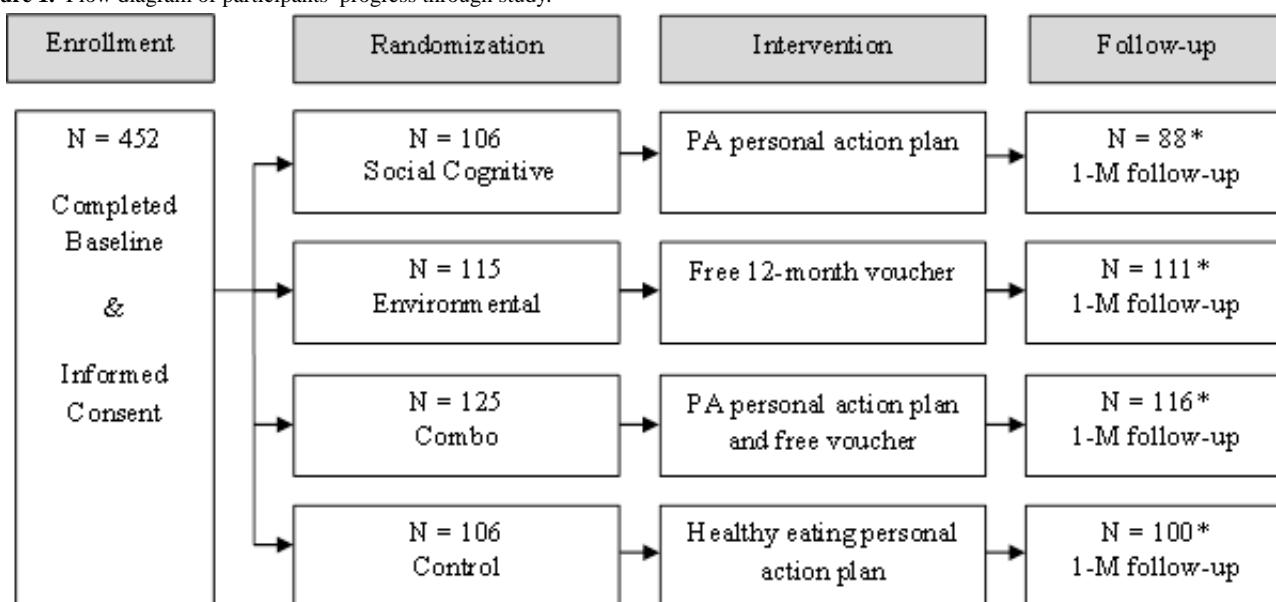
Changes in Physical Activity: Weekly Minutes of Moderate and Vigorous Physical Activity

Table 2 shows that participants from three study groups (SC, ENV, and COMBO) increased their level of PA at 1 month follow-up, while participants from the CON group reported a small decrease. A repeated-measures analysis with all participants found a main effect for time indicating that PA increased across the study sample from baseline to follow-up ($F_{1,441}$=30.03, P<.001). A time by condition interaction ($F_{3,441}$=8.33, P<.001) followed by post hoc analyses indicated that SC participants exhibited a significant increase in weekly PA participation (mean 45.1, SD 10.2) compared to CON (mean -2.5, SD 10.8, P=.004), and ENV (mean 8.3, SD 8.1, P<.05). Additionally, COMBO participants exhibited a significant increase in weekly PA participation (mean 53.4, SD 8.9) compared to CON (mean -2.5, SD 10.8, P<.001), and ENV (mean 8.3, SD 8.1, P=.003) participants. There were no differences between ENV and CON or between SC and COMBO.

Table 2. Study outcomes by condition and time.

Weekly minutes of moderate and vigorous PA	Environmental (n=113)	Social cognitive (n=105)	Combination (n=124)	Control (n=103)	Overall (n=445)
Baseline, mean (SE)	65.76 (7.04)	63.38 (7.30)	56.25 (6.72)	80.49 (7.37)	66.47 (3.55)
1 month, mean (SE)	74.04 (8.37)	108.46 (8.69)	109.71 (7.99)	77.97 (8.77)	92.55 (4.23)
Difference	+8.28[a]	+45.08[b]	+53.46[c]	-2.52[a,b,c]	+26.07
Meet PA public health recommendations, 1 month, %	15.9	27.6[d]	29.0[d]	15.5	22.2

[a]Denotes no difference between ENV and CON groups.

[b]Denotes difference between SC and CON groups: P=.004.

[c]Denotes difference between COMBO and CON groups: P<.001.

[d]Significantly different from control group, χ^2_3=10.343, P=.016.

Changes in Physical Activity: Meeting Physical Activity Public Health Recommendations

Table 2 depicts the percentage of participants in each study group who achieved public health recommendations. Since no participants met public health recommendations at baseline, these percentages represent total increase from baseline to 1-month follow-up. At 1-month follow-up, 15.9%, 27.6%, 29.0%, and 15.5% of the participants for the ENV, SC, COMBO, and CON groups, respectively, met the PA recommendations for aerobic activities. Chi-square test results indicate that SC and COMBO groups had significantly higher prevalence of meeting recommendations than CON group, χ^2_3=10.343, P=.016, while ENV participants did not significantly differ from CON participants.

Binary logistical regression analysis with all participants showed that SC (P=.036) and COMBO (P=.017) participants were significantly more likely to meet public health PA recommendations when compared to CON participants. ENV participants showed no significant difference from CON participants on meeting PA recommendations. Regression coefficients are shown in Table 3. Odds ratio results suggest that participants in the SC group (OR 2.075, 95% CI 1.047-4.110) and COMBO group (OR 2.224, 95% CI 1.151-4.301) were 108% and 122%, respectively, more likely to achieve current public health recommendations for aerobic exercise than their CON group counterparts. There were no differences between ENV group participants and CON group.

Table 3. Likelihood of meeting PA recommendation results (control group set as reference category).

Study group	B	SE	Wald	Sig.	OR	95% CI for exp. B
Environmental (n=113)	.030	0.374	0.006	0.937	1.030	0.495-2.145
Social cognitive (n=105)	.730	0.349	4.380	0.036	2.075	1.047-4.110
Combination (n=124)	.800	0.336	5.650	0.017	2.224	1.151-4.301

Discussion

Principal Findings

This study sought to determine whether a brief, tailored, computer-based, interactive personal action planning session operationalized using the 5 As could initiate increased PA participation among a group of sedentary cardiac patients. We hypothesized that participants who completed the computer-based interactive personal action planning session would exhibit significant increases in PA participation compared to control participants. This hypothesis was supported. We found that participants assigned to the SC and COMBO study arms experienced significant increases in PA at 1 month compared to CON participants who decreased their PA participation. This is a particularly important finding given that the current evidence on the effectiveness of tailored, technology-based PA programs is mixed at best [12,30-32]. In fact, Becker et al suggested that it would be unreasonable to expect PA change from low-dosage

interventions [46]. Our findings suggest the contrary, showing that a brief, low-dosage, tailored, computer-based interactive personal action planning session can in fact lead to PA behavior initiation among previously sedentary adults.

An ancillary aim of this study was to determine whether environmental or social-cognitive mechanisms to PA behavior change may impact PA initiation differentially. We found that participants who were exposed to the social-cognitive approach (SC and COMBO) exhibited significantly greater PA increases at 1 month compared to participants in the control group. On the other hand, participants who were exposed only to the environmental condition (ENV) showed no differences from the control condition. Together these findings suggest that a social cognitive approach to initiating PA may be a superior approach than only providing environmental access to PA through a complementary fitness facility membership.

Nevertheless, even though SC participants experienced the greatest increase in PA during the study period, participants in this arm also had a higher rate of non-completion of IVR calls. This was an unexpected but interesting finding. One explanation may be that non-completion was higher for SC participants due to the demand of the SC intervention relative to the ENV intervention. On the other hand, participants in the COMBO arm had a lower rate of non-completion even though they were also exposed to the SC content. Because the only intervention difference between the SC and COMBO arms was that the latter was exposed to the ENV content, an alternative explanation may be that the fitness facility access provided an incentive for participants to remain in the study and therefore increased the completion rate. These findings appear to suggest that while an individual approach may have a greater impact on PA initiation, environmental factors may also play a key role in the ongoing engagement of participants. As such, the combination of the environmental intervention with the personal action planning approach appears to lead to both increased rates of PA participation and better IVR completion rates overall.

Strengths and Limitations

This study was not without limitations. First, PA was measured using self-reported minutes of PA, leaving room for reporting error and/or bias. However, we suspect that participants may have provided more honest estimates of their PA participation since they were simply typing these numbers into the telephone keypad rather than reporting them directly to a research staff member, potentially reducing response bias [47,48]. In particular, self-administered questionnaires are thought to be

completed in a more relaxed and honest manner without the pressure of an external interviewer [47]. Our IVR approach may have provided participants with a more relaxed experience without the presence of the interviewer and with the convenience of completing the questionnaire on their own time at a location of their choosing (participants were allowed to select dates and times for IVR completion that were the most convenient for them). Nevertheless, the effect sizes observed in the current study must be interpreted with caution given the inherent danger of overestimation of physical activity behavior when using self-report measures. Second, change in PA was reported over a relatively short timeframe of 1 month. Although we observed significant increases in PA at 1 month, it is unclear whether this increase in PA participation would be maintained over a longer period of time. Future studies should examine the impact of an interactive computer intervention on long-term PA participation.

One of the strengths of this study was the low resources-high reach mode of intervention delivery. The interactive computer approach allowed for a tailored intervention that required minimal resources for delivery within a health care setting. Furthermore, the IVR contact at 1 month was tailored based on the participant's responses to the baseline interactive computer session and allowed for collection of additional data all while using minimal resources. Results from this study support the concept of a minimum intervention needed for change [49]. The interactive computer intervention used in this study was resource efficient and resulted in significant behavior change over a short period. As such, because of the feasibility of our interactive computer intervention delivered within the ETT system, this approach has potential for rapid and broad adoption and offers promise for integration within health care settings.

Conclusions

In general, our findings indicate that a social cognitive approach to PA behavior change may lead to a greater increase in PA than simply providing access to a fitness facility. However, this approach alone seemed to lead to higher IVR non-completion rates. While a single exposure to an interactive computer session was sufficient to lead to PA initiation, this approach together with providing access to a fitness facility appeared to be the best combination in order to initiate PA and increase IVR completion rates. These findings provide initial evidence that brief, low-dosage, tailored, interactive technology-based interventions to initiate PA can be feasibly delivered within the health care setting, in particular as part of the ETT system, and lead to increased PA among sedentary cardiac patients.

Conflicts of Interest

None declared.

References

1. Williamson J, Pahor M. Evidence regarding the benefits of physical exercise. Arch Intern Med 2010 Jan 25;170(2):124-125 [FREE Full text] [doi: 10.1001/archinternmed.2009.491] [Medline: 20101006]

2. Warburton DER, Nicol CW, Bredin SSD. Health benefits of physical activity: the evidence. CMAJ 2006 Mar 14;174(6):801-809 [FREE Full text] [doi: 10.1503/cmaj.051351] [Medline: 16534088]

3. Balady GJ, Chaitman B, Driscoll D, Foster C, Froelicher E, Gordon N, et al. Recommendations for cardiovascular screening, staffing, and emergency policies at health/fitness facilities. Circulation 1998 Jun 9;97(22):2283-2293 [FREE Full text] [Medline: 9631884]

4. Nelson ME, Rejeski WJ, Blair SN, Duncan PW, Judge JO, King AC, et al. Physical activity and public health in older adults: recommendation from the American College of Sports Medicine and the American Heart Association. Med Sci Sports Exerc 2007 Aug;39(8):1435-1445. [doi: 10.1249/mss.0b013e3180616aa2] [Medline: 17762378]

5. Carlson SA, Fulton JE, Schoenborn CA, Loustalot F. Trend and prevalence estimates based on the 2008 Physical Activity Guidelines for Americans. Am J Prev Med 2010 Oct;39(4):305-313. [doi: 10.1016/j.amepre.2010.06.006] [Medline: 20837280]

6. Leon AS, Franklin BA, Costa F, Balady GJ, Berra KA, Stewart KJ, Council on Clinical Cardiology (Subcommittee on Exercise, Cardiac Rehabilitation,Prevention), Council on Nutrition, Physical Activity, Metabolism (Subcommittee on Physical Activity), American association of Cardiovascular and Pulmonary Rehabilitation. Cardiac rehabilitation and secondary prevention of coronary heart disease: an American Heart Association scientific statement from the Council on Clinical Cardiology (Subcommittee on Exercise, Cardiac Rehabilitation, and Prevention) and the Council on Nutrition, Physical Activity, and Metabolism (Subcommittee on Physical Activity), in collaboration with the American association of Cardiovascular and Pulmonary Rehabilitation. Circulation 2005 Jan 25;111(3):369-376 [FREE Full text] [doi: 10.1161/01.CIR.0000151788.08740.5C] [Medline: 15668354]

7. Bandura A. Health promotion from the perspective of social cognitive theory. Psychology and Health 1998;13(4):623-649. [doi: 10.1080/088704498084074]

8. Conn VS, Hafdahl AR, Mehr DR. Interventions to increase physical activity among healthy adults: meta-analysis of outcomes. Am J Public Health 2011 Apr;101(4):751-758 [FREE Full text] [doi: 10.2105/AJPH.2010.194381] [Medline: 21330590]

9. Spence JC, Lee RE. Toward a comprehensive model of physical activity. Psychology of Sport and Exercise 2003 Jan;4(1):7-24. [doi: 10.1016/S1469-0292(02)00014-6]

10. Sallis J, Owen N, Fisher E. Ecological models of health behavior. In: Health behavior and health education: Theory, research, and practice. San Francisco: Jossey-Bass; 2008.

11. Dzewaltowski D, Noble J, Shaw J. Physical activity participation: Social cognitive theory versus the theories of reasoned action and planned behavior. Journal of Sport & Exercise Psychology 1990:388-405.

12. Lin JS, O'Connor E, Whitlock EP, Beil TL. Behavioral counseling to promote physical activity and a healthful diet to prevent cardiovascular disease in adults: a systematic review for the U.S. Preventive Services Task Force. Ann Intern Med 2010 Dec 7;153(11):736-750. [doi: 10.7326/0003-4819-153-11-201012070-00007] [Medline: 21135297]

13. Damush TM, Stewart AL, Mills KM, King AC, Ritter PL. Prevalence and correlates of physician recommendations to exercise among older adults. J Gerontol A Biol Sci Med Sci 1999 Aug;54(8):M423-M427. [Medline: 10496548]

14. Eakin EG, Glasgow RE. Recruitment of managed care Medicare patients for a physical activity study. Am J Health Promot 1997;12(2):98-101. [Medline: 10174668]

15. Glasgow RE, Eakin EG, Fisher EB, Bacak SJ, Brownson RC. Physician advice and support for physical activity: results from a national survey. Am J Prev Med 2001 Oct;21(3):189-196. [Medline: 11567839]

16. Wee CC. Physical activity counseling in primary care: the challenge of effecting behavioral change. JAMA 2001 Aug 8;286(6):717-719. [Medline: 11495623]

17. Hillsdon M, Foster C, Thorogood M. Interventions for promoting physical activity. Cochrane Database Syst Rev 2005(1):CD003180 [FREE Full text] [doi: 10.1002/14651858.CD003180.pub2] [Medline: 15674903]

18. Eakin E, Reeves M, Lawler S, Graves N, Oldenburg B, Del MC, et al. Telephone counseling for physical activity and diet in primary care patients. Am J Prev Med 2009 Feb;36(2):142-149. [doi: 10.1016/j.amepre.2008.09.042] [Medline: 19062240]

19. The Writing Group for the Activity Counseling Trial Research Group. Effects of Physical Activity Counseling in Primary Care: The Activity Counseling Trial: A Randomized Controlled Trial. JAMA 2001;286(6):677-687. [doi: 10.1001/jama.286.6.677]

20. Bennett GG, Glasgow RE. The delivery of public health interventions via the Internet: actualizing their potential. Annu Rev Public Health 2009;30:273-292. [doi: 10.1146/annurev.publhealth.031308.100235] [Medline: 19296777]

XSL·FO

RenderX

21. Dallery J, Riley W, Nahum-Shani I. Research designs to developevaluate technology-based health behavior interventions. In: Marsch L, Lord S, Dallery J, editors. Behavioral health care and technology: Using science-based innovations to transform practice. Carry, NC: Oxford University Press. , 978019931. Cary, NC: Oxford University Press; 2014:179-185.

22. Fanning J, Mullen SP, McAuley E. Increasing physical activity with mobile devices: a meta-analysis. J Med Internet Res 2012;14(6):e161 [FREE Full text] [doi: 10.2196/jmir.2171] [Medline: 23171838]

23. King AC, Ahn DK, Oliveira BM, Atienza AA, Castro CM, Gardner CD. Promoting physical activity through hand-held computer technology. Am J Prev Med 2008 Feb;34(2):138-142 [FREE Full text] [doi: 10.1016/j.amepre.2007.09.025] [Medline: 18201644]

24. Duncan TE, Duncan SC, Beauchamp N, Wells J, Ary DV. Development and evaluation of an interactive CD-ROM refusal skills program to prevent youth substance use: "refuse to use". J Behav Med 2000 Feb;23(1):59-72. [Medline: 10749011]

25. Block G, Wakimoto P, Metz D, Fujii ML, Feldman N, Mandel R, et al. A randomized trial of the Little by Little CD-ROM: demonstrated effectiveness in increasing fruit and vegetable intake in a low-income population. Prev Chronic Dis 2004 Jul;1(3):A08 [FREE Full text] [Medline: 15670429]

26. Casazza K, Ciccazzo M. Improving the dietary patterns of adolescents using a computer-based approach. J Sch Health 2006 Feb;76(2):43-46. [doi: 10.1111/j.1746-1561.2006.00067.x] [Medline: 16466465]

27. Tessaro I, Rye S, Parker L, Mangone C, McCrone S. Effectiveness of a nutrition intervention with rural low-income women. Am J Health Behav 2007;31(1):35-43. [doi: 10.5555/ajhb.2007.31.1.35] [Medline: 17181460]

28. Glasgow RE, Bull SS. Making a Difference With Interactive Technology: Considerations in Using and Evaluating Computerized Aids for Diabetes Self-Management Education. Diabetes Spectrum 2001 Apr 01;14(2):99-106. [doi: 10.2337/diaspect.14.2.99]

29. Parekh S, King D, Boyle FM, Vandelanotte C. Randomized controlled trial of a computer-tailored multiple health behaviour intervention in general practice: 12-month follow-up results. Int J Behav Nutr Phys Act 2014;11(1):41 [FREE Full text] [doi: 10.1186/1479-5868-11-41] [Medline: 24646165]

30. Kroeze W, Werkman A, Brug J. A systematic review of randomized trials on the effectiveness of computer-tailored education on physical activity and dietary behaviors. Ann Behav Med 2006 Jun;31(3):205-223. [doi: 10.1207/s15324796abm3103_2] [Medline: 16700634]

31. Norman GJ, Zabinski MF, Adams MA, Rosenberg DE, Yaroch AL, Atienza AA. A review of eHealth interventions for physical activity and dietary behavior change. Am J Prev Med 2007 Oct;33(4):336-345 [FREE Full text] [doi: 10.1016/j.amepre.2007.05.007] [Medline: 17888860]

32. van den Berg MH, Schoones JW, Vliet Vlieland TPM. Internet-based physical activity interventions: a systematic review of the literature. J Med Internet Res 2007;9(3):e26 [FREE Full text] [doi: 10.2196/jmir.9.3.e26] [Medline: 17942388]

33. Compernolle S, Vandelanotte C, Cardon G, De BI, De CK. Effectiveness of a web-based, computer-tailored, pedometer-based physical activity intervention for adults: a cluster randomized controlled trial. J Med Internet Res 2015;17(2):e38 [FREE Full text] [doi: 10.2196/jmir.3402] [Medline: 25665498]

34. Mouton A, Cloes M. Web-based interventions to promote physical activity by older adults: promising perspectives for a public health challenge. Arch Public Health 2013;71(1):16 [FREE Full text] [doi: 10.1186/0778-7367-71-16] [Medline: 23819885]

35. Brough C, Boyce S, Houchen-Wolloff L, Sewell L, Singh S. Evaluating the interactive web-based program, activate your heart, for cardiac rehabilitation patients: a pilot study. J Med Internet Res 2014;16(10):e242 [FREE Full text] [doi: 10.2196/jmir.3027] [Medline: 25359204]

36. Mouton A, Cloes M. Efficacy of a web-based, center-based or combined physical activity intervention among older adults. Health Educ Res 2015 Mar 15;30(3):422-435. [doi: 10.1093/her/cyv012] [Medline: 25772974]

37. Gibbons RJ, Balady GJ, Beasley JW, Bricker JT, Duvernoy WF, Froelicher VF, et al. ACC/AHA guidelines for exercise testing: executive summary. A report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines (Committee on Exercise Testing). Circulation 1997 Jul 1;96(1):345-354 [FREE Full text] [Medline: 9236456]

38. Borresen J, Lambert MI. Autonomic control of heart rate during and after exercise : measurements and implications for monitoring training status. Sports Med 2008;38(8):633-646. [Medline: 18620464]

39. Peels DA, Hoogenveen RR, Feenstra TL, Golsteijn RH, Bolman C, Mudde AN, et al. Long-term health outcomes and cost-effectiveness of a computer-tailored physical activity intervention among people aged over fifty: modelling the results of a randomized controlled trial. BMC Public Health 2014;14:1099 [FREE Full text] [doi: 10.1186/1471-2458-14-1099] [Medline: 25342517]

40. Estabrooks PA, Glasgow RE. Translating effective clinic-based physical activity interventions into practice. Am J Prev Med 2006 Oct;31(4 Suppl):S45-S56. [doi: 10.1016/j.amepre.2006.06.019] [Medline: 16979469]

41. Estabrooks PA, Glasgow RE, Xu S, Dzewaltowski DA, Lee RE, Thomas D, et al. Building a multiple modality, theory-based physical activity intervention: The development of CardiACTION!. Psychol Sport Exerc 2011 Jan 1;12(1):46-53 [FREE Full text] [doi: 10.1016/j.psychsport.2010.04.012] [Medline: 22473308]

42. Estabrooks PA, Glasgow RE, Dzewaltowski DA. Physical activity promotion through primary care. JAMA 2003 Jun 11;289(22):2913-2916. [doi: 10.1001/jama.289.22.2913] [Medline: 12799388]

XSL·FO
RenderX

43. Macera CA, Pratt M. Public health surveillance of physical activity. Res Q Exerc Sport 2000 Jun;71(2 Suppl):S97-103. [Medline: 10925831]

44. Brownson RC, Jones DA, Pratt M, Blanton C, Heath GW. Measuring physical activity with the behavioral risk factor surveillance system. Med Sci Sports Exerc 2000 Nov;32(11):1913-1918. [Medline: 11079522]

45. Centers for Disease Control and Prevention (CDC). Prevalence of physical activity, including lifestyle activities among adults--United States, 2000-2001. MMWR Morb Mortal Wkly Rep 2003 Aug 15;52(32):764-769 [FREE Full text] [Medline: 12917582]

46. Becker A, Herzberg D, Marsden N, Thomanek S, Jung H, Leonhardt C. A new computer-based counselling system for the promotion of physical activity in patients with chronic diseases--results from a pilot study. Patient Educ Couns 2011 May;83(2):195-202. [doi: 10.1016/j.pec.2010.05.024] [Medline: 20573467]

47. Siemiatycki J. A comparison of mail, telephone, and home interview strategies for household health surveys. Am J Public Health 1979 Mar;69(3):238-245. [Medline: 420369]

48. McHorney CA, Kosinski M, Ware JE. Comparisons of the costs and quality of norms for the SF-36 health survey collected by mail versus telephone interview: results from a national survey. Med Care 1994 Jun;32(6):551-567. [Medline: 8189774]

49. Glasgow RE, Fisher L, Strycker LA, Hessler D, Toobert DJ, King DK, et al. Minimal intervention needed for change: definition, use, and value for improving health and health research. Transl Behav Med 2014 Mar;4(1):26-33 [FREE Full text] [doi: 10.1007/s13142-013-0232-1] [Medline: 24653774]

Abbreviations

5 As: Ask, Advise, Assess, Assist, and Arrange
AHA: American Heart Association
CHD: coronary heart disease
COMBO: combined PA intervention arm of both ENV and SC approaches
CON: matched contact nutrition control arm
ENV: increased environmental accessibility to PA resources arm
ETT: exercise treadmill testing
GIS: geographic information system
IVR: interactive voice response system
PA: physical activity
SC: tailored social cognitive PA intervention arm
USPTF: US Preventive Task Force

Review

Mobile Apps for Bipolar Disorder: A Systematic Review of Features and Content Quality

Jennifer Nicholas[1,2], BSc(Hons) BA; Mark Erik Larsen[1], DPhil; Judith Proudfoot[1], PhD; Helen Christensen[1], PhD

[1]Black Dog Institute, University of New South Wales, Sydney, Australia

[2]School of Psychiatry, Faculty of Medicine, University of New South Wales, Sydney, Australia

Corresponding Author:
Jennifer Nicholas, BSc(Hons) BA
Black Dog Institute
University of New South Wales
Hospital Road, Prince of Wales Hospital
Randwick
Sydney, 2031
Australia
Phone: 61 293829260
Fax: 61 293828208
Email: j.nicholas@blackdog.org.au

Abstract

Background: With continued increases in smartphone ownership, researchers and clinicians are investigating the use of this technology to enhance the management of chronic illnesses such as bipolar disorder (BD). Smartphones can be used to deliver interventions and psychoeducation, supplement treatment, and enhance therapeutic reach in BD, as apps are cost-effective, accessible, anonymous, and convenient. While the evidence-based development of BD apps is in its infancy, there has been an explosion of publicly available apps. However, the opportunity for mHealth to assist in the self-management of BD is only feasible if apps are of appropriate quality.

Objective: Our aim was to identify the types of apps currently available for BD in the Google Play and iOS stores and to assess their features and the quality of their content.

Methods: A systematic review framework was applied to the search, screening, and assessment of apps. We searched the Australian Google Play and iOS stores for English-language apps developed for people with BD. The comprehensiveness and quality of information was assessed against core psychoeducation principles and current BD treatment guidelines. Management tools were evaluated with reference to the best-practice resources for the specific area. General app features, and privacy and security were also assessed.

Results: Of the 571 apps identified, 82 were included in the review. Of these, 32 apps provided information and the remaining 50 were management tools including screening and assessment (n=10), symptom monitoring (n=35), community support (n=4), and treatment (n=1). Not even a quarter of apps (18/82, 22%) addressed privacy and security by providing a privacy policy. Overall, apps providing information covered a third (4/11, 36%) of the core psychoeducation principles and even fewer (2/13, 15%) best-practice guidelines. Only a third (10/32, 31%) cited their information source. Neither comprehensiveness of psychoeducation information (r=-.11, P=.80) nor adherence to best-practice guidelines (r=-.02, P=.96) were significantly correlated with average user ratings. Symptom monitoring apps generally failed to monitor critical information such as medication (20/35, 57%) and sleep (18/35, 51%), and the majority of self-assessment apps did not use validated screening measures (6/10, 60%).

Conclusions: In general, the content of currently available apps for BD is not in line with practice guidelines or established self-management principles. Apps also fail to provide important information to help users assess their quality, with most lacking source citation and a privacy policy. Therefore, both consumers and clinicians should exercise caution with app selection. While mHealth offers great opportunities for the development of quality evidence-based mobile interventions, new frameworks for mobile mental health research are needed to ensure the timely availability of evidence-based apps to the public.

Introduction

Symptom monitoring is a fundamental component of the self-management of bipolar disorder (BD) that benefits from being performed in real-time and under natural conditions [1]. With mobile phone penetration reaching 97% worldwide, and smartphones accounting for the majority of phone sales in 2014 [2], researchers and clinicians are investigating the use of smartphone technology to enhance the management of chronic illnesses such as BD. As smartphones are personal devices that are carried on the user, mobile apps are a perfect platform for self-management. Smartphones can also be used to deliver interventions and psychoeducation, supplement treatment, and enhance therapeutic reach in BD, as apps are cost-effective, accessible, anonymous, and convenient.

A study by Proudfoot et al [3] found that consumers hold positive attitudes towards evidence-based mental health apps, with 76% of survey responders indicating that they would be interested in using their smartphones for monitoring and self-management of mental health. This acceptability has also been shown in mobile programs for BD, with personal digital assistant [4,5] and short message service based interventions [1] proving effective and acceptable to consumers. Similar programs are now being developed for smartphones. For instance, Wenze et al [6] are developing an expanded smartphone version of their successful PDA-based "Increasing Adherence in Bipolar Disorder" intervention, which will incorporate psychoeducation and Cognitive Behavioral Therapy (CBT) sessions using smartphones alongside in-person treatment to address medication adherence. A smartphone app based on Interpersonal and Social Rhythms Therapy (IPSRT) augmented with objective data from phone sensors is also in production. This app, called "MoodRhythm", aims to stabilize daily routines and rhythms to prevent episode-onset [7]. Also investigating the potential use of objective data alongside self-report, the monitoring, treatment, and prediction of bipolar disorder episodes (MONARCA) study app attempts to validate objective indicators of affect change by investigating correlations with user-inputted and clinician mood ratings [8]. Additionally, Grünerbl et al [9] are investigating extending this research to enable the app to identify mood state changes based solely on objective sensor data. However, to our knowledge, there are currently no apps available for people with BD that have been subject to research evaluation to determine effectiveness.

Previous reviews have found that apps currently available to consumers for other health conditions are unlikely to be supported by research data, nor be developed with reference to evidence-based practice. For example, an investigation of the content of smoking cessation apps found that apps had low concordance with guidelines, with an average guideline adherence score of 7.8 out of a possible total of 60 [10]. Examining other chronic conditions with a strong self-management focus, a review of apps for asthma self-management determined that only 5% of those providing information were comprehensive, while 44% made

recommendations that were not in line with treatment guidelines [11]. Similarly, a review of apps for human immunodeficiency virus / sexually transmitted disease prevention and care found that only 11% of assessed apps covered all literature-driven prevention areas [12]. Moreover, Donker et al [13] performed a systematic review of apps for mental health and noted the striking disparity between the number of apps available and the number of scientific articles assessing the evidence base of mental health apps. Similarly, a simultaneous review of health-related apps and research literature for a number of conditions noted the contrast between the number of research publications regarding the use of mobile technology to manage the investigated illnesses and the profusion of available apps [14].

An examination of the Google Play and Apple iOS app stores indicates that there is an abundance of publically available apps for BD despite the limited published research. Therefore, there is little information accessible to consumers about the quality of available apps beyond the user "star" rating system and user reviews. However, the popularity of an app is not necessarily a reliable indicator of its quality, effectiveness, or evidence base. The opportunity for mHealth to assist in the self-management of BD is feasible only if apps are of appropriate quality, and therefore, the aim of the current study was to identify the types of self-management apps available in the Google Play and iOS app stores and to assess the comprehensiveness and quality of their content for BD.

Methods

Search Criteria and Selection

A systematic review framework was applied to the search, screening, and assessment of apps.

Search terms to identify apps developed specifically for BD were derived through a preliminary search of the Google Play store. Relevant synonyms and layperson alternatives to the identified terms were determined and also included in the search. Layperson alternatives were included because our aim was to assess apps that were specifically designed for consumers with BD, a group that may not use the technical terms of the disorder. In this way, the following search terms were identified: bipolar, bi-polar, "manic depression", "mood swings", "mania" AND "mood", cyclothymia, and cyclothymic. On July 16, 2014, these terms were used to identify publically accessible apps for BD in the Australian Google Play and iOS app stores. Android apps were identified by a search of the Google Play store via the Web interface, while iOS apps were identified via the iTunes/iOS search application programming interface.

Details of apps identified from the app stores were extracted from the search results including app name, description, and price. Duplicate apps that were retrieved by multiple search terms on the same device platform were removed. Conversely, apps with different versions (eg, free/paid) or which appeared across different platforms were retained for comparison. Two

reviewers (JN and ML) independently screened the app store description of each app. Apps were included in the review if the following criteria were met: (1) developed for BD, (2) aimed at consumers and/or caregivers, (3) available for download through official Android/iOS app stores, and (4) available in English. Discrepancies were identified and discussed until consensus was achieved.

App Assessment

Apps meeting the inclusion criteria were downloaded onto either a Samsung Galaxy S4 mini (Android version 4.2.2) or an iPhone 5s (iOS version 7.1) for complete assessment. Two reviewers (JN and ML) performed the assessment of apps using a standard data extraction form.

Descriptive characteristics related to the following features were extracted: (1) app accessibility, including platform, price, need for network connectivity, number of downloads, and average user-scored star rating for the app's entire history, (2) primary function, such as app store classification, app function (see below), and target audience, (3) app source, including provider type and details, crisis information, disclaimer presence, and app update schedule, and (4) privacy and security, presence of an accessible privacy policy, and the ability to password-protect data.

The function of each app was determined through use of the app and classified as providing either information or self-management tools. Additional assessment was performed based on the identified function. Apps containing information about BD were assessed according to the comprehensiveness of its psychoeducation information and the quality of the information, as reflected by its concordance with evidence-based guidelines. The comprehensiveness of BD psychoeducational information was assessed according to 11 fundamental topics of face-to-face psychoeducation interventions [15]. Core statements were derived from the main topics and goals of psychoeducation with reference to Colom and Vieta's Psychoeducational Manual for Bipolar Disorder [15] (Table 1). A clinical psychologist with experience with BD then reviewed the statements for accuracy and completeness.

Table 1. Core components of psychoeducation for BD.[a]

Topic	Criteria
1. Facts about the nature of BD.	States that BD is biological in nature but interacts with environmental factors (diathesis-stress model).
	States that BD is chronic and recurrent in nature and has a cyclic course.
2. Information on common symptoms of each phase of the disorder.	States common symptoms of both (hypo)mania and depression.
	States that risk of suicide is associated with BD.
3. Treatment options for each illness phase.	Outlines available pharmacotherapy for each illness phase: depression, mania, and prophylaxis.
	Mentions psychotherapy as a treatment option for BD.
4. Treatment adherence, withdrawal, and side effects.	States the importance of treatment adherence and states that risk of episode relapse is associated with abandonment of treatment.
5. Substance use in BD.	States that psychoactive substances may trigger episodes.
6. Identification of episode warning signs (EWS).	States common EWS of (hypo)mania and depression.
	States that EWS vary between people and indicates the importance of identifying personal episode warning signs.
7. Support networks and the role of support people or caregivers.	Describes a support person as someone who is close to the patient, aware of their BD, and knowledgeable about the disorder.
	States that a support network can assist in early detection of episodes.
8. The role of an action plan.	States the importance of having an action plan that provides a guide to stay well when episode EWS are detected.
	States common strategies that help prevent episodes once EWS are detected.
9. The importance of routine.	States that regular habits, including sleep, are of importance in BD.
	States that regular schedules and better structuring of activities are key in BD management.
10. Information on stress management and problem solving.	States that stress plays an important role in episode relapse.
	States there are tools that help manage stress and anxiety.
11. Episode risk-factors/triggers.	States common external factors that contribute to episode relapse.

[a]Topics based on Colom and Vieta's psychoeducation for BD manual [15].

The quality of BD information was assessed as the degree of concordance with the most recent treatment guidelines [16-19] and prominent meta-analyses and reviews of evidence-based treatment for the acute phases of the disorder [20,21]. Thirteen statements were derived from treatment guidelines and meta-analyses for use as quality indicators in the app review

(Table 2; [16-24]). A psychiatrist previously involved in the development of the Australian and New Zealand Clinical Practice Guidelines for the Treatment of Bipolar Disorder [22] then assessed and adjusted the statements for accuracy and comprehensiveness. Comprehensiveness and quality of information were calculated as a score out of 11 and 13 respectively, indicating how many of the core psychoeducation and guideline-derived statements were covered by the app's information.

Apps providing common tools for the management of BD were assessed against best-practice resources for their purpose. Screening and assessment apps were assessed according to whether the app used a validated BD screening questionnaire or how closely it reflected current diagnostic criteria [25,26]. Symptom monitoring apps were evaluated according to their similarity to well-known paper-and-pencil monitoring tools [27,28]. Treatment apps for BD were assessed against statements derived from the core goals of three commonly used therapies for the disorder: CBT [29], IPSRT [30], and Family-Focused Therapy (FFT) [31]. As with statements used to assess the comprehensiveness and quality of information, treatment statements were reviewed by a clinical psychologist with experience with BD. Community support apps, which provided access to BD-orientated discussion boards, were assessed based on their methods of communication, and the level of moderation or oversight.

Table 2. App quality assessment statements derived from BD treatment guidelines and meta-analyses.

Statement	Associated guideline
1. Initiation of an atypical antipsychotic and/or mood stabilizer for the treatment of acute mania.	"Efficacy of lithium and divalproex is well established…substantial RCT data support atypical antipsychotic monotherapy with olanzapine, risperidone ER, quetiapine, ziprasidone, and aripiprazole for the first-line treatment of acute mania" Yatham et al, 2013, pp 4, 6 [19]
	"Overall, risperidone, olanzapine, and haloperidol seem to be the most effective evidence-based options for the treatment of manic episodes" Cipriani et al, 2011, pp 1314 [20]
2. Use of an atypical antipsychotic or mood stabilizer, with or without an antidepressant, for the treatment of bipolar depression.	"Lithium, lamotrigine, quetiapine…monotherapies, as well as lithium or divalproex plus selective serotonin reuptake inhibitor, olanzapine plus SSRI…recommended as first-line choices for bipolar depression" Yatham et al, 2013, pp 9 [19]
	"Adjunctive antidepressants may be used for an acute bipolar I or II depressive episode when there is a history of previous positive response to antidepressants" Pacchiarotti et al, 2013, pp 1253 [21]
3. Antidepressant subtypes tricyclic antidepressants and serotonin-norepinephrine reuptake inhibitors (SNRIs) are more likely to cause switching than serotonin-specific reuptake inhibitors (SSRIs).	"The risk of mood switching is considered to be…somewhat greater with tri-and tetracyclics (and perhaps some SNRIs) than with most modern antidepressants" Pacchiarotti et al, 2013, pp 1256 [21]
	"monotherapy with some antidepressants, especially tricyclics, without an accompanying mood stabilizer, however, may be associated with an increased rate of treatment emergent affective switches (TEAS)" Grunze et al, 2010, pp 92 [17]
4. Lithium, an atypical antipsychotic, or lamotrigine (where depression predominates) for maintenance treatment of BD.	"Lithium, divalproex, olanzapine, and quetiapine, as well as lamotrigine (primarily for preventing depression)…continue to be first-line monotherapy options for maintenance treatment of BD" Yatham et al, 2013, pp 14 [19]
	"Lithium, olanzapine or valproate should be considered for long-term treatment of bipolar disorder" NCCMH, 2006, pp 5 [23]
5. Change monotherapy or use combination therapy for treatment resistance.	"No response after 2 weeks, switch to another first choice medication, in severe mania, consider combination" Grunze et al, 2009, pp 104 [16]
	"If the patient has frequent relapses, or symptoms continue to cause functional impairment, switching to an alternative monotherapy or adding a second prophylactic agent should be considered." NCCMH, 2006, pp 5 [23]
6. The use of electroconvulsive therapy (ECT) for treatment resistant acute symptoms (particularly depression, but also mania).	"Especially in very severe and psychotic depression, or in depression with severe psychomotor retardation, ECT has a major role" Grunze et al, 2010, pp 100 [17]
	"ECT is recommended for bipolar depression after an antidepressant trial has failed" RANZCP, 2004, pp 288 [22]
	"ECT is still a valuable last resource in severe delirious mania which is otherwise treatment refractory" Grunze et al, 2009, pp 102 [16]
7. Careful monitoring of blood levels is required where those correlate with treatment response (eg, lithium, valproate).	"Plasma concentrations need to be checked on a frequent and regular basis until equilibrium in the therapeutic range as been achieved and thereafter. It is recommended to check every 3-6 months" Grunze et al, 2013, pp 186 [18]
	"lithium is up titrated in small steps guided by individual experience and plasma level monitoring" Grunze et al, 2013, pp 186 [18]
8. Careful monitoring of potential physical complications or side effects of treatments is required (eg, kidney, thyroid, and calcium with lithium; glucose and lipids with antipsychotics).	"Renal and thyroid function should also be checked regularly, every 6-12 months depending on risks" Grunze et al, 2013, pp 186 [18]
	"Complete medication and laboratory investigations should be performed at baseline, with ongoing monitoring for weight changes and adverse effects of medication" Yatham et al, 2013, pp 29 [19]
9. Women informed about ensuring that their medications are safe to take during breastfeeding and pregnancy.	"Important that women with bipolar disorder receive education early in the course of illness about the effects of mood stabilizing and other medications on contraceptive effectiveness, as well as the need to plan medication management during pregnancy and the postpartum period" Yatham et al, 2005, pp 33 [24]
10. Seek medical professional advice and/or a second opinion in diagnosis of BD in children, due to the controversy in this area.	"The presentation and diagnosis of BD in children and adolescents remains controversial…diagnostic criteria for BD may not be systematically applied in some clinical settings." Yatham et al, 2013, pp 19 [19]

XSL•FO
RenderX

Statement	Associated guideline
11. Outlines difficulties in the treatment of rapid cycling BD.	"Rapid cycling…is associated with greater severity of illness on a number of clinical measures" Yatham et al, 2005, pp 30 [24]
	"The prophylactic use of lithium in rapid cycling patients has been discouraged for a long time based on the observation of insufficient acute and prophylactic efficacy in these patients" Grunze et al, 2013, pp 184 [18]
12. Optimal treatment for most patients with BD will include psychological treatment as well as medication.	"When used as adjuncts to pharmacotherapy, psychosocial interventions…have demonstrated significant benefits, both in the treatment of acute depressive episodes and also as long-term maintenance treatment…providing psychological treatments—and, in particular, brief psychoeducation, which has been demonstrated to be as effective as CBT at much lower cost—is an essential aspect of managing patients with BD" Yatham et al, 2013, pp 4 [19]
	"The primary long-term treatments are pharmacological, but psychological and psychosocial interventions have an important part to play" NCCMH, 2006, pp 33 [23]
13. Indicates that most patients benefit considerably from treatment for their BD.	"The advent of these therapies, both drug and psychological, means that the majority of patients with this recurrent and disabling condition may be effectively treated" RANZCP, 2004, pp 299 [22]

Data Analysis

Descriptive statistics were used to summarize the results of the app assessment. Pearson product-moment correlation was performed to examine relationships between comprehensiveness and quality of information, and the average user rating of an app. Mann-Whitney U tests were performed to examine differences in comprehensiveness and quality of information by app price. Statistical significance was set at $P<.05$.

Results

Google Play and iOS app store searches identified 571 potential apps, of which 133 were removed as duplicates (identified on the same platform through more than one search term). Of the remaining 438 apps, 82 met inclusion criteria (see Multimedia Appendix 1 for a list of included apps). Details about the inclusion and exclusion of apps are provided in Figure 1.

Figure 1. Flow diagram illustrating the exclusion of apps at various stages of the study.
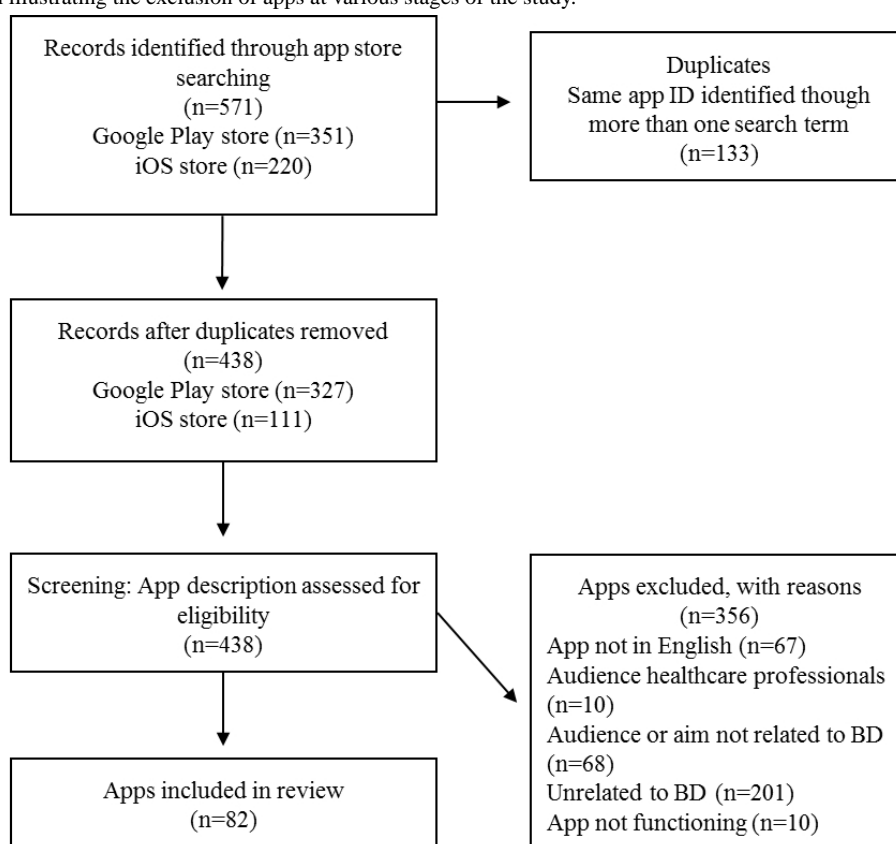
Description

Accessibility

Of the 82 included apps, 59 were available on the Android platform, 23 were available on iOS devices, and ten had versions for both platforms. Examining costs, 49 apps were free, and the median cost of paid apps was AU$1.70, with a minimum price of AU$.99 and a maximum of AU$16.99 (mean AU$3.05). Over half (n=48) did not require an active Internet connection after installation, 20 would not function in the absence of an Internet connection, and 14 needed connectivity for some features. Only the Google Play store reported the number of downloads of each app to date, indicated by a range. The median number of downloads was between 1000 and 5000 (information available for 53 of the apps). Fourteen of these 53 apps had been installed less than 100 times. The app store descriptions of 28 apps indicated that an insufficient number of users had rated the app over its lifetime to provide an average user rating. The mean average user rating of the remaining 54 apps was 3.5 out of a possible 5.

Primary Function

On the app stores, apps for BD were classified as health and fitness (n=48), medical (n=15), lifestyle (n=6), books (n=3), social (n=2), and entertainment (n=1), with genre information unavailable for seven apps. Thirty-one apps provided written information about the disorder, and one used animation to raise awareness. The remaining 50 apps provided tools for the management of BD and included ten screening and assessment tests, 35 symptom monitoring apps, four community support discussion boards, and one treatment app. No apps performed multiple functions.

For 78 apps, the function of the app was clearly stated in the app store description. Apps with unclear descriptions either introduced the disorder without explaining the role of the app, were unclear between two distinct functions, or provided information for an unrelated app. Forty-nine apps were developed for use by patients, with four apps specifically designed for caregivers, 14 for the general public, and 15 for both patients and caregivers. According to the app store descriptions, no apps were developed for a specific BD clinical subgroup.

App Source

As far as we can discern, none of the 82 apps assessed had been subject to research evaluation. No apps cited published material about the feasibility or effectiveness of the app, either within the app or the app store description. Furthermore, there was a lack of Google Scholar results upon a search of app names.

The majority of apps were developed by commercial providers (n=48) or private individuals (n=20), followed by institutions (n=3), other types of providers (n=6) including, clinical groups and not-for-profit organizations, and unspecified (n=5). The app provider's name was included in 46 apps, of which 30 also included contact details; eight apps included only provider contact details. A disclaimer outlining that the app was not a replacement for consultation with a medical professional was provided in 23 of the 82 apps, including seven screening and assessment apps, and eight information apps. A statement providing information and contact details of crisis resources was provided in only nine apps for BD. A quarter of apps (n=19) had been updated within the last 6 months, with half of the apps updated within the last 17 months.

Privacy and Confidentiality

Privacy policies were available, either in the app or as a link from the app store description, for 18 of the 82 apps. Of the 50 apps able to record personal data (screening and assessment, symptom monitoring, community support, and treatment apps), 12 had a privacy policy. An account or password was needed to access these data for 19 of these apps.

Information

Overview

Apps containing information were more readily available on the Android (n=29) compared to the iOS (n=3) platform. Of the 32 apps that provided health information, seven presented general health news stories about a variety of conditions, including BD, and a further six provided no disorder-specific information to assess. The remaining 19 information apps were assessed for comprehensiveness and quality.

Of these 19 apps, all contained in-app information content, and four linked directly to additional external information. In all cases, external material was referenced; however, just four credited the source of their in-app information. Although only three of the apps were classified on the app stores as books, the assessment revealed 12 information apps were repurposed ebooks. Although unattributed, many shared the same ebook content, with nine apps having the same information content as another app, but with differences in user interface layout, connectivity reliance, or wording.

Comprehensiveness of Bipolar Disorder Psychoeducational Information

The comprehensiveness of the psychoeducation information presented in the apps is reported below (Table 3). Overall, apps covered an average of four of the 11 psychoeducation statements (SD 3.0). Two statements, concerning the importance of treatment adherence and the development of action plans as a guide to stay well, were not addressed by any app. Five of the apps contained no BD specific psychoeducation information in line with best practice, while the most comprehensive apps covered seven statements (seven apps). These seven apps were reproductions of the same content, although uncited.

Table 3. Comprehensiveness of psychoeducation topics covered by BD information apps.

Topic	Apps covering topic, n (%)
1. Facts about the nature of BD.	11 (58)
2. Information on common symptoms of each phase of the disorder.	14 (74)
3. Treatment options for each illness phase.	5 (26)
4. Treatment adherence, withdrawal, and side effects.	0 (0)
5. Substance use in BD.	8 (42)
6. Identification of episode early warning signs.	7 (37)
7. Support networks and the role of support people or caregivers.	8 (42)
8. The role of an action plan.	0 (0)
9. The importance of routine.	8 (42)
10. Information on stress management and problem solving.	8 (42)
11. Episode risk-factors/triggers.	3 (16)

Quality of BD information: Concordance With Evidence-Based Practice

Of the 13 evidence-based statements extracted from treatment guidelines for BD, an average of two statements were covered by apps (SD 2.3; Table 4). Only three apps provided information aligned with more than three guidelines, covering four, seven, and eight statements, and six apps addressed no statements. The app that covered seven statements comprised 50 chapters, with no logical structure and repeated unreferenced information. In contrast, the app that addressed eight statements contained information structured into six concise sections and referenced NHS Choices [32].

Table 4. Quality of BD app information: concordance to BD treatment guidelines.

Statement	Apps covering topic, n (%)
1. Initiation of an atypical antipsychotic and/or mood stabilizer for the treatment of acute mania.	3 (16)
2. The use of an atypical antipsychotic or mood stabilizer, with or without an antidepressant, for the treatment of bipolar depression.	2 (11)
3. Antidepressant subtypes tricyclic antidepressants and SNRIs are more likely to cause switching than SSRIs.	0 (0)
4. Lithium, an atypical antipsychotic, or lamotrigine (where depression predominates) for maintenance treatment of BD.	4 (21)
5. Change monotherapy or use combination therapy for treatment resistance.	1 (5)
6. The use of ECT for treatment resistant acute symptoms (particularly depression, but also mania).	9 (47)
7. Careful monitoring of blood levels is required where those correlate with treatment response (eg, lithium, valproate).	2 (11)
8. Careful monitoring of potential physical complications or side effects of treatments is required (eg, kidney, thyroid, and calcium with lithium; glucose and lipids with antipsychotics).	1 (5)
9. Women informed about ensuring that their medications are safe to take during breastfeeding and pregnancy.	1 (5)
10. Seek medical professional advice and/or a second opinion in diagnosis of BD in children, due to the controversy in this area.	1 (5)
11. Outlines difficulties in the treatment of rapid cycling BD.	1 (5)
12. Optimal treatment for most patients with BD will include psychological treatment as well as medication.	10 (53)
13. Most patients benefit considerably from treatment for their BD.	10 (53)

Across the apps that provided information about BD, four contained incorrect information, three of which incorrectly differentiated the different types of BD. Two contained critically wrong information; the first suggested, "take a shot of hard liquor a [sic] hour before bed" (app ID 404) to assist with sleep during a manic episode. The second incorrectly informed users about the types of BD and indicated that BD was contagious "sometimes [BD] can transfer to another relative if they spend too much time with you and listen to your depressive life" (app ID 28).

User Ratings

Neither comprehensiveness of psychoeducation information (r=-.11, P=.80) nor information quality (r=-.02, P=.96) were significantly correlated with average user ratings. A Mann-Whitney U test also revealed no significant difference in

comprehensiveness (U=30.5, Z=-1.2, P=.22) or quality (U=42.5, Z=-.21, P=.83) between free and paid apps.

Tools

Screening and Assessment

Ten apps offered self-assessments to screen for BD, of which three were for children. Only three cited the source of the test, two of which used a validated screening measure (the M3 checklist) [33]. Two additional apps used the Mood Disorder Questionnaire [34] without attribution. In total, only four of the 10 screening apps used validated measures, and six of the tests screened for only manic symptoms.

Of the apps that asked about symptoms of depression, three had a duty-of-care message at the end of the test, referring the user to clinical support if questions about suicidal ideation were answered positively. Upon informing the user they had screened positively for BD, eight apps recommended visiting a health care professional, and two had the option for users to share the result on social media.

Symptom Monitoring

In total, 35 apps were symptom monitoring tools, which aim to assist users with tracking the symptoms of their BD. The number of apps monitoring mood, medication, sleep, and other symptoms is shown in Table 5. A mood scale designed for BD was defined as a continuous scale with "depressed" and "manic" at the extremes. Other scales used to monitor mood included using different emoticons to represent emotions (n=10), rating mood on a generic scale (n=8), or had users select from or rate a list of moods (n=4). Thirty apps monitored factors not commonly included on established mood charts, with energy levels and anxiety the most common additions. These tracking options were customizable in nine apps, allowing users to monitor factors specifically relevant to their mood states. General customization of other features such as esthetics, profiles, and data reports was possible in 21 apps.

Table 5. Symptoms monitored by symptom monitoring apps for BD.

Factors monitored	n (%)
Mood	34 (97)
Mood on scale designed for BD	12 (34)
Medication	15 (43)
Sleep	17 (49)
Functioning	3 (9)
Section for free-notes	32 (91)
Mixed episodes or mood switches	2 (6)
Menstruation	5 (14)
Other	30 (86)

Alerts reminding users to track their mood were available in 22 of the 35 monitoring apps. Of these, 13 allowed users to designate a reminder time, six asked users to designate how many times per day to be reminded, and three had options for both. Over a sampling period of 3 days, seven apps failed to notify the user as directed.

None of the 35 monitoring apps had a duty-of-care alert, that is, a message suggesting users contact a health care provider. This was tested by logging 3 consecutive days of severely depressed mood and indicating suicidal ideation in free-text notes. Inputted monitoring data was presented graphically by 30 apps, could be exported from 20, and shared via social media by seven. Back-up of personal monitoring data was available in only six apps. No app used the sensor capabilities of the smartphone to enhance monitoring data beyond what is possible with paper-and-pencil resources, or as an attempt to validate subjective mood reporting with objective data.

Treatment

There was only one treatment app, which delivered a CBT intervention, although it had not been specifically developed for BD. An analysis of treatment content against common treatment features of CBT, IPSRT, and FFT confirmed that the app was based on CBT principles, with three of the six CBT items endorsed, and none related to other therapies. However, the source or evidence base of the CBT presented in the app was not referenced.

Community Support

Four community apps gave users access to BD-orientated discussion boards, where members sought information and support. App communities were for consumers and family/caregivers and allowed communication via forum posts and private messages. Site owners for each app monitored forum post communication.

Discussion

Principle Findings and Comparison With Prior Work

This assessment of apps for BD indicates that most currently available apps do not reference clinical practice guidelines, standard psychoeducation information, or established self-management tools. Apps were available for a variety of uses including disorder information (38%), and management tools such as screening and assessment (12%), symptom monitoring (43%), and community support (5%). Interestingly, no apps combined these functions to provide both information and self-management tools. Furthermore, while all the assessed

apps mentioned BD in their app store description, the content evaluation indicated few had been designed specifically for the disorder. For example, only a third of mood monitoring apps provided a BD specific mood scale, with the remaining apps using scales inadequate for the population. Furthermore, merely half of the apps claiming to provide information about BD actually provided disorder-specific information to be assessed for comprehensiveness and quality.

Disappointingly, apps that did provide BD specific information had low levels of adherence to quality assessment criteria, neither comprehensively addressing main psychoeducation domains, nor endorsing evidence-based practice guidelines. The app that correctly addressed the most treatment guideline criteria was Your MD's Symptom Checker (app ID 39), which referenced NHS Choices [32], and presented information concisely and coherently (only BD information was assessed). Overall, comprehensiveness and quality scores were low and not correlated with average user ratings, confirming that an app's popularity is not an accurate gauge of its content. However, with only a third of information apps citing their information source, users are denied other important information by which to make informed judgments about an app's robustness and credibility.

Apps that offered tools to assist with the management of BD were also inconsistent in quality. Few screening and assessment apps cited the source of the test; additionally, a Google search of each app's assessment questions revealed that less than half used validated screening measures. Only one treatment app was included in the review, and while it covered half of the CBT treatment domains, it was not specifically designed for the needs of people with BD. Symptom monitoring was the most prevalent self-management tool offered by apps. However, only a third allowed users to track all three symptoms standard in traditional monitoring tools: mood, sleep, and medication. The high proportion of monitoring apps that did not feature a notification or alert function to remind users to input daily mood data was unexpected. As the ability to remind users to complete monitoring in situ is a major advantage of this technology, it is surprising that more than a third of monitoring apps did not utilize this function, and a further seven did not ensure the feature functioned as intended. It was also an unexpected finding that no symptom monitoring apps used the smartphone's sensors to supplement the user-inputted monitoring data. Torous et al [35] found that all published articles regarding apps for bipolar disorder focused on the use of passive sensor-enabled data collection with supplemental or no user input. This is another example of the disconnect between emerging research in this area, and apps currently available for the disorder.

These results are consistent with findings of reviews assessing apps of other heath domains. Unfortunately, many of these reviews examine quality using the information provided in the app store description rather than assessing app content against evidence-based clinical guidelines [36]. However, those that do assess content have found similarly low endorsement of quality criteria. Huckvale et al [11] assessed asthma information against evidence-based asthma treatment guidelines and reported that the most endorsed guideline was cited by only 32% of apps. A similarly low concordance with guidelines was found in a review

of smoking cessation apps [10]. While a recent review of depression apps did not assess content, it highlighted another difficulty faced by users of health apps. Shen et al [37] reported that a search for depression on the five major app stores yielded three times more non-related apps than depression apps [37]. In the current study, 78% of apps were excluded due to not being related to BD or not having a BD specific aim. This suggests an app's quality is not the only challenge consumers face in identifying appropriate apps, as irrelevant alternatives dominate search results. Together, these findings suggest that the results of the current study reflect a wider problem of app availability and quality across health conditions.

Rarely considered, the lack of privacy policies made available to app users was another area in which apps developed for BD fell short of optimal standards. A recent review by Dehling et al of information security and privacy of mobile health apps characterized apps that had access to mental health information as high sensitivity [38]. The review identified health monitors, state of health tests, treatment reminders, and health records, which included disease management tools, as the highest privacy and security risks [38]. However, only a quarter of the assessed BD self-management tools, which would fall into the above categories, provided a statement indicating how data would be protected, stored, and shared. Sunyaev et al [39] reported a similar absence of mHealth privacy policy availability. Their assessment of health apps found that only 30.5% provided privacy policies, two-thirds of which did not address the app specifically [39]. This represents a problematic lack of transparent reporting of how apps handle personal and mental health-related data. As the use of apps in health care becomes more common, the topic of information privacy will continue to grow in relevance, as privacy and security consistently feature as major considerations of users [3,38]. Therefore, the time to ensure structures are in place to address privacy and security is now.

While apps asked users to input personal health data, very few responded to indications that users were unwell. In fact, only three apps responded to users indicating severe extremes of mood or suicidal ideation. A lack of response to severity of recorded symptoms was also noted by Huckvale et al [11] in asthma management tools. Although both psychoeducation and clinical guidelines emphasize the importance of the ability to identify and respond to indications of mood change, few apps assisted users by informing them of such changes. This identification and response to indicators of change in mood forms the basis of action plans or stay well plans. Action plans are a common feature of psychoeducation [15] and combine information about early warning signs and factors that may ameliorate them into a plan that aims to prevent episode relapse. Action plans are also a common self-management tool used by people with BD to stay well [40,41]. A cohort of individuals that had not experienced an episode of BD for more than 2 years were interviewed about their strategies for staying well [41]. All participants discussed the importance of a tailored and revised stay well plan, which identified triggers and early warning signs and designated strategies to intervene. However, no information apps mentioned action plans, and no self-management tools assisted users in the development of an

action plan. Together with the lack of response to user input, this represents a missed opportunity for the facilitation of important self-management practices by mobile technology.

Limitations

The current study is not without its limitations. A possible limitation is that the apps assessed were those available through the Australian Google Play and iOS app stores, and app availability may differ globally. This may affect the generalizability of results to apps available in other locations. However, a search of the Australian, British, and American iTunes app stores using the primary search term "bipolar" indicated that results were remarkably comparable. In total, 97% of apps were available in all three regions, with only three apps exclusively available in the American store. Therefore, at least in English-speaking countries, apps appear to be universally available, and this review provides a representative overview of the apps available internationally. A second possible limitation is that while this study is novel in that it provides, to our knowledge, the first assessment of apps for BD in respect to clinical guidelines, psychoeducation principles, and self-management tools, it does so for a snapshot of apps available when the search was performed in July 2014. As such,

with changing nature of the Google Play and iOS app stores, the marketplace of apps for BD will differ by time of publication.

Conclusions

With the exponential increase in the use of smartphones, there has been an increased interest in the use of apps for information dissemination and disease management. However, this study's findings highlight important shortcomings in current app marketplace offerings for BD, suggesting that apps are developed independently of research data, and without reference to best practice clinical guidelines. These results indicate that clinicians looking to recommend apps to supplement treatment should exercise caution with app selection and that policy makers and the research community need to consider ways of assuring app quality. There is an opportunity for mental health research to develop quality evidence-based mobile interventions that assist with the management of BD. However, with the domains of research and technology currently moving at different paces, new frameworks for mobile mental health research are needed to prevent a lag in availability and to ensure that evidence-based apps are available to consumers.

Conflicts of Interest

None declared.

Multimedia Appendix 1

List of apps included in the review.

[PDF File (Adobe PDF File), 33KB - jmir_v17i8e198_app1.pdf]

References

1. Bopp JM, Miklowitz DJ, Goodwin GM, Stevens W, Rendell JM, Geddes JR. The longitudinal course of bipolar disorder as revealed through weekly text messaging: a feasibility study. Bipolar Disord 2010 May;12(3):327-334 [FREE Full text] [doi: 10.1111/j.1399-5618.2010.00807.x] [Medline: 20565440]
2. Ericsson Mobility Report: On the pulse of the networked society. 2015. URL: http://www.ericsson.com/res/docs/2015/ericsson-mobility-report-feb-2015-interim.pdf [accessed 2015-04-26] [WebCite Cache ID 6Y4KpDV9G]
3. Proudfoot J, Parker G, Hadzi Pavlovic D, Manicavasagar V, Adler E, Whitton A. Community attitudes to the appropriation of mobile phones for monitoring and managing depression, anxiety, and stress. J Med Internet Res 2010;12(5):e64 [FREE Full text] [doi: 10.2196/jmir.1475] [Medline: 21169174]
4. Depp CA, Mausbach B, Granholm E, Cardenas V, Ben-Zeev D, Patterson TL, et al. Mobile interventions for severe mental illness: design and preliminary data from three approaches. J Nerv Ment Dis 2010 Oct;198(10):715-721 [FREE Full text] [doi: 10.1097/NMD.0b013e3181f49ea3] [Medline: 20921861]
5. Wenze SJ, Armey MF, Miller IW. Feasibility and Acceptability of a Mobile Intervention to Improve Treatment Adherence in Bipolar Disorder: A Pilot Study. Behav Modif 2014 Jul;38(4):497-515 [FREE Full text] [doi: 10.1177/0145445513518421] [Medline: 24402464]

XSL·FO

RenderX

6.   Wenze S, Miller IW, Armey M. An ecological momentary intervention to improve treatment adherence in bipolar disorder: Feasibility, acceptability, and preliminary data. Bipolar Disorders 2013;15(Suppl. 1):149.

7.   Frank E, Matthews M, Choudhury T, Voida S, Abdullah S. Developing a smart phone app to monitor mood, social rhythms, sleep and social activity: Technology to support effective management of bipolar disorder. 2013 Presented at: The American College of Neuropsychopharmacology Annual Conference; Dec. 8-12, 2013; Florida, United States.

8.   Faurholt-Jepsen M, Frost M, Vinberg M, Christensen EM, Bardram JE, Kessing LV. Smartphone data as objective measures of bipolar disorder symptoms. Psychiatry Res 2014 Jun 30;217(1-2):124-127. [doi: 10.1016/j.psychres.2014.03.009] [Medline: 24679993]

9.   Grünerbl A, Muaremi A, Osmani V, Bahle G, Ohler S, Tröster G, et al. Smartphone-based recognition of states and state changes in bipolar disorder patients. IEEE J Biomed Health Inform 2015 Jan;19(1):140-148. [doi: 10.1109/JBHI.2014.2343154] [Medline: 25073181]

10.   Abroms LC, Padmanabhan N, Thaweethai L, Phillips T. iPhone apps for smoking cessation: a content analysis. Am J Prev Med 2011 Mar;40(3):279-285 [FREE Full text] [doi: 10.1016/j.amepre.2010.10.032] [Medline: 21335258]

11.   Huckvale K, Car M, Morrison C, Car J. Apps for asthma self-management: a systematic assessment of content and tools. BMC Med 2012;10:144 [FREE Full text] [doi: 10.1186/1741-7015-10-144] [Medline: 23171675]

12.   Muessig KE, Pike EC, Legrand S, Hightow-Weidman LB. Mobile phone applications for the care and prevention of HIV and other sexually transmitted diseases: a review. J Med Internet Res 2013;15(1):e1 [FREE Full text] [doi: 10.2196/jmir.2301] [Medline: 23291245]

13.   Donker T, Petrie K, Proudfoot J, Clarke J, Birch M, Christensen H. Smartphones for smarter delivery of mental health programs: a systematic review. J Med Internet Res 2013;15(11):e247 [FREE Full text] [doi: 10.2196/jmir.2791] [Medline: 24240579]

14.   Martínez-Pérez B, de la Torre-Díez I, López-Coronado M. Mobile health applications for the most prevalent conditions by the World Health Organization: review and analysis. J Med Internet Res 2013;15(6):e120 [FREE Full text] [doi: 10.2196/jmir.2600] [Medline: 23770578]

15.   Colom F, Vieta E. Psychoeducation manual for bipolar disorder. Melbourne: Cambridge University Press; 2006.

16.   Grunze H, Vieta E, Goodwin GM, Bowden C, Licht RW, Möller HJ, et al. The World Federation of Societies of Biological Psychiatry (WFSBP) guidelines for the biological treatment of bipolar disorders: update 2009 on the treatment of acute mania. World J Biol Psychiatry 2009;10(2):85-116. [doi: 10.1080/15622970902823202] [Medline: 19347775]

17.   Grunze H, Vieta E, Goodwin GM, Bowden C, Licht RW, Möller HJ, et al; WFSBP Task Force On Treatment Guidelines For Bipolar Disorders. The World Federation of Societies of Biological Psychiatry (WFSBP) Guidelines for the Biological Treatment of Bipolar Disorders: Update 2010 on the treatment of acute bipolar depression. World J Biol Psychiatry 2010 Mar;11(2):81-109. [doi: 10.3109/15622970903555881] [Medline: 20148751]

18.   Grunze H, Vieta E, Goodwin GM, Bowden C, Licht RW, Möller HJ, et al; WFSBP Task Force on Treatment Guidelines for Bipolar Disorders. The World Federation of Societies of Biological Psychiatry (WFSBP) guidelines for the biological treatment of bipolar disorders: update 2012 on the long-term treatment of bipolar disorder. World J Biol Psychiatry 2013 Apr;14(3):154-219. [doi: 10.3109/15622975.2013.770551] [Medline: 23480132]

19.   Yatham LN, Kennedy SH, Parikh SV, Schaffer A, Beaulieu S, Alda M, et al. Canadian Network for Mood and Anxiety Treatments (CANMAT) and International Society for Bipolar Disorders (ISBD) collaborative update of CANMAT guidelines for the management of patients with bipolar disorder: update 2013. Bipolar Disord 2013 Feb;15(1):1-44. [doi: 10.1111/bdi.12025] [Medline: 23237061]

20.   Cipriani A, Barbui C, Salanti G, Rendell J, Brown R, Stockton S, et al. Comparative efficacy and acceptability of antimanic drugs in acute mania: a multiple-treatments meta-analysis. Lancet 2011 Oct 8;378(9799):1306-1315. [doi: 10.1016/S0140-6736(11)60873-8] [Medline: 21851976]

21.   Pacchiarotti I, Bond DJ, Baldessarini RJ, Nolen WA, Grunze H, Licht RW, et al. The International Society for Bipolar Disorders (ISBD) task force report on antidepressant use in bipolar disorders. Am J Psychiatry 2013 Nov;170(11):1249-1262 [FREE Full text] [doi: 10.1176/appi.ajp.2013.13020185] [Medline: 24030475]

22.   Royal AustralianNew Zealand College of Psychiatrists Clinical Practice Guidelines Team for Bipolar Disorder. Australian and New Zealand clinical practice guidelines for the treatment of bipolar disorder. Aust N Z J Psychiatry 2004 May;38(5):280-305. [doi: 10.1111/j.1440-1614.2004.01356.x] [Medline: 15144505]

23.   National Collaborating Centre for Mental Health. Bipolar Disorder: the Management of Bipolar Disorder in Adults, Children and Adolescents, in Primary and Secondary Care. Leicester: The British Psychological Society; 2006.

24.   Yatham LN, Kennedy SH, O'Donovan C, Parikh S, MacQueen G, McIntyre R, et al; Canadian Network for Mood and Anxiety Treatments. Canadian Network for Mood and Anxiety Treatments (CANMAT) guidelines for the management of patients with bipolar disorder: consensus and controversies. Bipolar Disord 2005;7 Suppl 3:5-69. [doi: 10.1111/j.1399-5618.2005.00219.x] [Medline: 15952957]

25.   American Psychiatric Association. Diagnostic and statistical manual of mental disorders. 5th edition. London: American Psychiatric Pub; 2013.

26.   World Health Organization. The ICD-10 classification of mental and behavioural disorders: clinical descriptions and diagnostic guidelines. Geneva: World Health Organization; 1992.

27. Black Dog Institute. Daily mood graph for bipolar disorder. 2014. URL: http://www.blackdoginstitute.org.au/docs/12. MoodChartforBipolarDisorder.pdf [accessed 2015-04-26] [WebCite Cache ID 6Y4KfUjmM]

28. Robins LN, Helzer JE, Croughan J, Ratcliff KS. National Institute of Mental Health Diagnostic Interview Schedule. Its history, characteristics, and validity. Arch Gen Psychiatry 1981 Apr;38(4):381-389. [Medline: 6260053]

29. Lam D, Jones S, Hayward P. Cognitive therapy for bipolar disorder: A therapist's guide to concepts, methods and practice. 2nd edition. Singapore: John Wiley & Sons; 2010.

30. Frank E. Treating bipolar disorder: A clinician's guide to interpersonal and social rhythm therapy. London: Guilford Press; 2007.

31. Miklowitz D. Bipolar disorder: A family-focused treatment approach. London: Guilford Press; 2010.

32. Department of Health. NHS Choices: Bipolar disorder. 2014. URL: http://www.nhs.uk/Conditions/Bipolar-disorder/Pages/ Introduction.aspx [accessed 2015-04-26] [WebCite Cache ID 6Y4LBDV0V]

33. Gaynes BN, DeVeaugh-Geiss J, Weir S, Gu H, MacPherson C, Schulberg HC, et al. Feasibility and diagnostic validity of the M-3 checklist: a brief, self-rated screen for depressive, bipolar, anxiety, and post-traumatic stress disorders in primary care. Ann Fam Med 2010;8(2):160-169 [FREE Full text] [doi: 10.1370/afm.1092] [Medline: 20212303]

34. Hirschfeld RM, Williams JB, Spitzer RL, Calabrese JR, Flynn L, Keck PE, et al. Development and validation of a screening instrument for bipolar spectrum disorder: the Mood Disorder Questionnaire. Am J Psychiatry 2000 Nov;157(11):1873-1875. [Medline: 11058490]

35. Torous J, Powell AC. Current research and trends in the use of smartphone applications for mood disorders. Internet Interventions 2015 May;2(2):169-173.

36. BinDhim NF, Hawkey A, Trevena L. A systematic review of quality assessment methods for smartphone health apps. Telemed J E Health 2015 Feb;21(2):97-104. [doi: 10.1089/tmj.2014.0088] [Medline: 25469795]

37. Shen N, Levitan M, Johnson A, Bender JL, Hamilton-Page M, Jadad AAR, et al. Finding a depression app: a review and content analysis of the depression app marketplace. JMIR Mhealth Uhealth 2015;3(1):e16 [FREE Full text] [doi: 10.2196/mhealth.3713] [Medline: 25689790]

38. Dehling T, Gao F, Schneider S, Sunyaev A. Exploring the Far Side of Mobile Health: Information Security and Privacy of Mobile Health Apps on iOS and Android. JMIR Mhealth Uhealth 2015;3(1):e8 [FREE Full text] [doi: 10.2196/mhealth.3672] [Medline: 25599627]

39. Sunyaev A, Dehling T, Taylor PL, Mandl KD. Availability and quality of mobile health app privacy policies. J Am Med Inform Assoc 2015 Apr;22(e1):e28-e33. [doi: 10.1136/amiajnl-2013-002605] [Medline: 25147247]

40. Murray G, Suto M, Hole R, Hale S, Amari E, Michalak EE. Self-management strategies used by 'high functioning' individuals with bipolar disorder: from research to clinical practice. Clin Psychol Psychother 2011;18(2):95-109. [doi: 10.1002/cpp.710] [Medline: 20572206]

41. Russell SJ, Browne JL. Staying well with bipolar disorder. Aust N Z J Psychiatry 2005 Mar;39(3):187-193. [doi: 10.1111/j.1440-1614.2005.01542.x] [Medline: 15701069]

Abbreviations

BD: bipolar disorder
CBT: cognitive behavioral therapy
ECT: electroconvulsive therapy
EWS: episode warning signs
IPSRT: interpersonal and social rhythms therapy
FFT: family-focused therapy
SNRI: serotonin-norepinephrine reuptake inhibitors
SSRI: serotonin-specific reuptake inhibitors

Original Paper

Successful Organizational Strategies to Sustain Use of A-CHESS: A Mobile Intervention for Individuals With Alcohol Use Disorders

James H Ford II[1], PhD; Esra Alagoz[2], PhD; Susan Dinauer[2], BA; Kimberly A Johnson[2], MBA, PhD; Klaren Pe-Romashko[2], MS; David H Gustafson[2], PhD

[1]University of Wisconsin - Madison, Center for Health Systems Research and Analysis, Madison, WI, United States

[2]University of Wisconsin - Madison, Center for Health Enhancement Systems Studies, Madison, WI, United States

Corresponding Author:
James H Ford II, PhD
University of Wisconsin - Madison
Center for Health Systems Research and Analysis
610 Walnut Street
Madison, WI, 53726
United States
Phone: 1 608 262 4748
Fax: 1 608 263 4523
Email: jhfordii@wisc.edu

Abstract

Background: Mobile health (mHealth) services are growing in importance in health care research with the advancement of wireless networks, tablets, and mobile phone technologies. These technologies offer a wide range of applications that cover the spectrum of health care delivery. Although preliminary experiments in mHealth demonstrate promising results, more robust real-world evidence is needed for widespread adoption and sustainment of these technologies.

Objective: Our aim was to identify the problems/challenges associated with sustained use of an mHealth addiction recovery support app and to determine strategies used by agencies that successfully sustained client use of A-CHESS.

Methods: Qualitative inquiry assessed staff perceptions about organizational attributes and strategies associated with sustained use of the mobile app, A-CHESS. A total of 73 interviews of clinicians and administrators were conducted. The initial interviews (n=36) occurred at the implementation of A-CHESS. Follow-up interviews (n=37) occurred approximately 12 and 24 months later. A coding scheme was developed and Multiuser NVivo was used to manage and analyze the blinded interview data.

Results: Successful strategies used by treatment providers to sustain A-CHESS included (1) strong leadership support, (2) use of client feedback reports to follow up on non-engaged clients, (3) identify passionate staff and incorporate A-CHESS discussions in weekly meetings, (4) develop A-CHESS guidelines related to client use, (5) establish internal work groups to engage clients, and (6) establish a financial strategy to sustain A-CHESS use. The study also identified attributes of A-CHESS that enhanced as well as inhibited its sustainability.

Conclusions: Mobile apps can play an important role in health care delivery. However, providers will need to develop strategies for engaging both staff and patients in ongoing use of the apps. They will also need to rework business processes to accommodate the changes in communication frequency and style, learn to use app data for decision making, and identify financing mechanisms for supporting these changes.

(J Med Internet Res 2015;17(8):e201)   doi:10.2196/jmir.3965

KEYWORDS
mHealth; substance abuse disorder; sustainability; funding; engagement; staff; client; implementation

Introduction

Emerging as a subsegment of eHealth technologies, mobile health (mHealth) services are growing in importance in health care with the advancement of wireless networks, tablets, and mobile phone technologies [1]. mHealth technologies offer a wide range of applications that cover the spectrum of health care delivery from measurement and diagnostics to treatment and recovery support. A 2013 Pew report states that 56% of the US population owned a smartphone in 2013, which is a 10%

increase from the previous year [2]. The possibilities for use of mHealth technology to address care delivery, patient monitoring, and adherence support have created substantial enthusiasm among developers, providers, policy makers, and researchers to use mHealth tools [3]. The Global Mobile Health Market report proposes that one-third of the smartphone users will use some kind of health care app in 2015 [4]. Among the 1980 available therapeutic apps, chronic disease management is the most common type of app [5]. Addiction recovery support apps fall into this category, with substance use disorders for some people being categorized as a chronic condition [6,7]. These apps offer personalized recovery monitoring and easy access to online support communities [8-16]. While some apps rely on text messaging to monitor alcohol and drug use and send reminders to users as an intervention [17,18], more current comprehensive recovery management systems that integrate mobile Internet capabilities and sensors like global positioning satellites (GPS) and accelerometers have also been introduced [9,19]. With personalized systems tailored specifically for user needs and 24/7 availability to patients and caregivers, comprehensive recovery management systems have the potential to take mHealth to a new level if they are well developed, adopted, and used by clients and supported by providers.

Although preliminary experiments in mHealth demonstrate promising results in data collection for health care research [20], medical and health care education [21], remote health care assistance in developing nations [22], and improved treatment outcomes [19], robust real-world evidence is needed for widespread adoption and sustainment of these technologies [23,24]. Drawbacks such as counselor resistance and perceptions of increased staff workload, as well as the cost of owning a smartphone or implementation of the technology, impede the widespread utilization of mHealth technologies in a sustainable manner [25,26].

The purpose of this research was to study the sustainability of the implementation of an mHealth app (A-CHESS) designed to support addiction recovery. A-CHESS is a mobile phone app designed to help people maintain recovery and prevent relapse from drug or alcohol addiction. In this study, the app was provided to the individual by their treatment agency and included ongoing online and sometimes in-person support from agency counselors. At a minimum, agencies provided connection to a 24-hour crisis line within the A-CHESS app for after-hours coverage.

A-CHESS is consistent with self-determination theory and Marlatt's stages-of-relapse model. The various A-CHESS services align with the elements of self-determination theory: competence, relatedness, and autonomy [27,28]. The app also incorporates a weekly check-in assessment using a version of the Brief Alcohol Monitor (BAM) survey to obtain patient data on recent alcohol and other drug use as well as status on protective and risk factors [29]. A-CHESS uses the BAM survey information to provide feedback for the individual and includes

a Bayesian predictive model that estimates the likelihood that the patient will relapse in the coming weeks [30]. Agency counselors are able to see the results of the surveys if the user agrees to share that information.

A-CHESS assists an individual in recovery through services in the following categories: (1) connection and communication with peers via personal profiles and "walls", discussion groups, and private messages, (2) assessment of relapse risks through daily check-ins, weekly surveys, and targeted feedback, (3) alerts and reminders to encourage sustained recovery, (4) addiction-related educational materials and recovery news, and (5) tools to manage stress and risky situations that could jeopardize recovery. Figure 1 shows images of the A-CHESS home screen, and Figure 2 shows the Panic Button options. Counselors use an A-CHESS tool to access information on how their clients are doing and connect with them via A-CHESS private messaging. Figure 3 shows a visual of the counselor home screen. In addition, counselors participate in discussion groups and are part of the support team for each individual.

The A-CHESS mobile phone app has been previously tested in a randomized trial involving 349 study participants and found to reduce risky drinking days and improve abstinence [19]. However, a clinical trial provides supports that are not available in most addiction treatment settings. Even though the intervention was efficacious, the question of whether such a technology could be implemented and sustained in a standard clinical practice remains unanswered.

While different frameworks exist to identify facilitators and barriers related to the implementation of organizational change, sustainability of organizational change is an emerging area of research. Existing sustainability frameworks suggest that sustainability is attributable to the (1) internal organizational supports, (2) contextual influences of the external environment, and (3) characteristics of the intervention [31-34]. Innovation characteristics include stakeholder relationships, innovation integrity, effectiveness, and quality of initial implementation efforts, and alignment between innovation and stakeholder (eg, agency staff or patients) needs. Internal organizational supports include champion roles, leadership, resources, policies and procedures, and expert support (ie, coaching). The environment context explores the effects of financial, regulatory, and policy changes. Rather than using a specific sustainability framework (eg, Conceptual Framework for Sustainability of Public Health Programs or Dynamic Sustainability Framework) [31-34], we examined the different elements from each framework to identify innovation, internal organizational, and external environmental factors influencing A-CHESS sustainability. The study sought to distinguish the different strategies employed between agencies that were sustainers and non-sustainers of A-CHESS and also to identify challenges associated with sustained use of A-CHESS beyond grant-funded randomized controlled trials or pilots.

Figure 1. A-CHESS main menu.
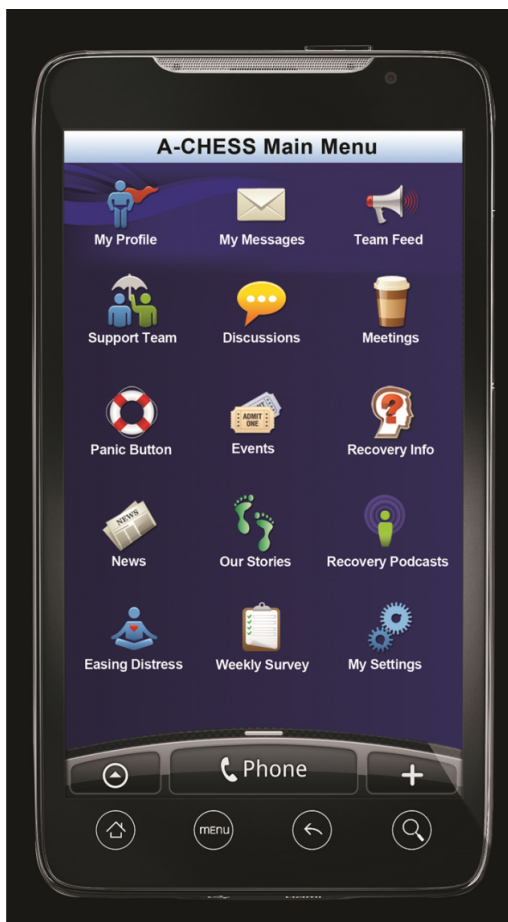
Figure 2. A-CHESS panic button options.
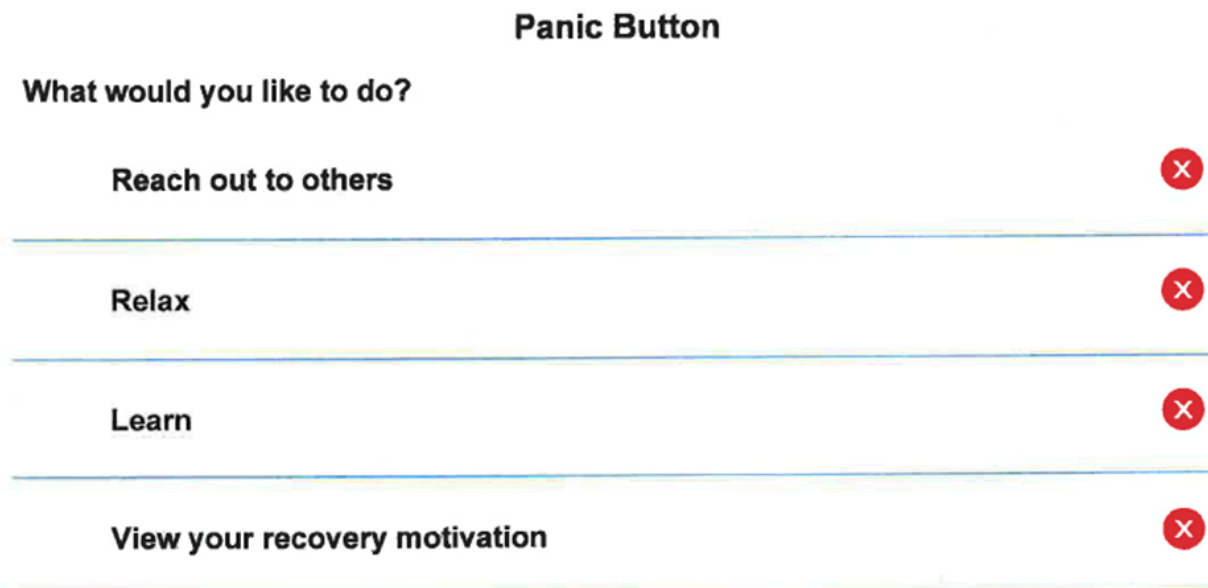
Figure 3. Screenshot of A-CHESS counselor home screen.

Methods

Setting

Agencies involved in this study were all members of the CHESS Health Education Consortium. The consortium brought together University of Wisconsin-Madison researchers and health care organizations to develop and research new health information technologies developed at the UW-Madison Center for Health Enhancement Systems Studies. The focus of the Consortium was to learn how to effectively implement A-CHESS in real-world settings. Treatment agencies that expressed interest in using A-CHESS in their clinics were notified of the opportunity to participate in the research Consortium. Notification took place via phone calls, emails, and at a conference for substance treatment professionals. As part of their Consortium membership, treatment agencies made a donation to the UW-Madison that funded Consortium activities. They also agreed to participate in studies that were generated from the interests of Consortium members. The consortium started in April 2011 and ended in December 2014.

The member agencies made the decision on how A-CHESS would be used within their organizations (eg, what client population, how many participants, what level of care). Members of the Consortium shared information on problems and best practices when implementing and using A-CHESS. The UW-Madison team adapted A-CHESS and expanded administrative tools to help with its implementation and utilization in the different treatment agencies. For example, agencies could add content, develop surveys, customize the meeting locator, or set up a drug testing schedule. The team also revised the use reports to improve ease to use by staff.

Interview Process

Organizations participating in the Consortium were interviewed at baseline using questions adapted from a previous study conducted with a drug court program using A-CHESS. The previous study was primarily interested in implementing rather than sustaining A-CHESS. The revised interview guide adapted the questions to focus on sustainable implementation and new questions were added to capture processes that would enhance sustainability (available on request from first author). Counselors using A-CHESS with clients were interviewed regarding their use of the app, their perceptions regarding any changes in workload, or how the app had/had not changed their interactions with clients. Program managers were asked about counselor attitudes, workflows, and supervision. The staff interview was repeated at 12 months and 24 months into the implementation process, using some of the questions from the baseline survey. These interviews also asked about organizational and innovation attributes (eg, ease of use or components that users like) that have been found in previous studies to affect sustainability of innovations [31-34]. CEOs/administrators of participating organizations were also interviewed. The baseline interview asked about cost, population, implementation, and thoughts on A-CHESS. The 12-month and 24-month interviews focused on environmental, organizational, and innovation attributes that might have influenced the sustainability of A-CHESS.

Recruitment

Agency senior leaders were informed of the study at the time they agreed to participate in the consortium. UW-Madison research staff contacted them via email to invite them to participate in an interview and also to identify other agency staff to interview about the A-CHESS implementation. The agency staff was sent separate individual emails that explained the focus of the interview, amount of time required, confidentiality issues, and that participation was voluntary. Consent with study participants was conducted at the start of the initial interview by the UW-Madison research study staff. The study received human subjects' approval from the UW-Madison Social and Behavioral Sciences Institutional Review Board (SE 2011-0568).

Data Collection

We conducted 36 baseline and 37 follow-up interviews of staff (counselors and administrators) from 14 agencies. Two UW staff (SD and KP) scheduled and conducted the interviews. Once completed, the interviews were de-identified to remove information that might indicate either a staff member's identity or the agency identity.

Definition of Sustainability

While multiple individuals were interviewed within a treatment provider, the presence or absence of strategies to support A-CHESS sustainability were summarized at the organizational level. The primary outcome measure is to assess short-term sustainability, which is defined as whether the agency continued to use A-CHESS (or another mHealth intervention) after the end of the Consortium in December 2014.

Analytical Approach

The coding scheme was constructed based on theories and frameworks related to both implementation and sustainability [33-35]. We also considered a list of a priori codes that emerged based on how the interview questions were constructed during the preliminary analyses of the dataset. The objective was to capture essential aspects that might influence or prevent sustainable implementation of A-CHESS. After finalizing the coding scheme, we piloted it on a subset of data and refined it through discussions among the researchers into nine main analytic themes with 42 codes nested within themes. Table 1 represents the full analytic framework used in our analysis.

To manage and analyze the blinded interview data, we used Multiuser NVivo, a qualitative data analysis software developed by QSR International. NVivo has components that allow multiple researchers to work collaboratively on data corpus. We utilized the "tree nodes" in NVivo as a structured, hierarchical placeholder for the coding scheme. These tree nodes were then issued to related sections of the dataset allowing the users to run queries on specific codes of interest. We used "annotations" and "memos" to mark the content that required further exploration from coders. These initialed notes were especially helpful for communicating ideas to other collaborators, as they were visible to anyone with access to the document.

A team of 4 researchers collaborated in coding the dataset in NVivo. Each researcher was assigned to a set that included the initial, 12-month, and 24-month interviews. We first completed coding 10% of the data and created comparison reports among coders using "coding comparison query", a function included in the software [36]. The interrater reliability among coders was 92%. We then allocated the remaining interviews among coders to proceed with the analysis.

The analysis focused on the short-term sustainability of A-CHESS at the organizational level. Input from staff and leadership provided different perspectives about the strategies employed. The interview guides asked similar questions of both staff and leadership, with the primary difference being that leadership was asked additional questions about how the external environment (eg, regulations or funding) might influence A-CHESS sustainability. The baseline interviews and follow-up interviews as well as interviews by staff type were initially analyzed separately to explore themes related to the implementation and sustainability of A-CHESS. However, as we moved further with our analysis, we found that the line between implementation and sustainability became blurred as many of our codes (such as "monitoring of sustainability" and "implementation barriers"), as well as the opinions of staff and leadership, overlapped. We realized that a complete picture of the organizational strategies employed to support A-CHESS sustainability could not be achieved without integrating the results into a holistic picture for the organization. Therefore, we did not draw a sharp line between the definitions of implementation and sustainability and did not distinguish between responses from staff and leadership.

Table 1. A-CHESS coding scheme.

Domain	Code		Description
Physical structure	Age		Age of the interviewee
	Gender		Gender of the interviewee
	Employment status		Administrator or staff, full time or part time.
	Weekly scheduled shifts		Hours of weekly scheduled shifts, including weekly night shifts and extra shifts.
	Client population		Any mention of client population, clients served with A-CHESS, any other characteristic that describes client norms, trends, or patterns, number of clients.
Communication	Method of communication		Any mention of how they communicate with clients, email, phone, text messaging, and preferred method of communication.
	Hours spent communicating		Any mention of hours spent communicating with clients, hours spent with one client, off-duty hours spent with clients.
	Client awareness		Any mention of knowledge of clients' health welfare, availability of emergency care, and clients' support mechanisms.
Readiness for A-CHESS	Access to technology		Any mention of technology they use regularly, use of electronic records, and access to Internet.
	Prior training		Any mention of attendance to webinars/conferences, whether they have continued training, whether it was sufficient, whether there were any helpful mediums, suggestions.
	Assumed benefits		Any mention of benefits of using A-CHESS with clients prior to the implementation of the tool.
	Reactions		Any mention of reactions, statements about learning curve prior to the implementation of the tool.
A-CHESS implementation	Implementation barriers	Technical barriers	Any mention of applications not working, difficult to use/understand applications.
		Resistance	Any mention of barriers to implementation such as staff resistance, leadership resistance, client resistance
		Insufficient funding	Any mention of payer restrictions, expenses, or cost indicated as implementation barrier.
		Other factors	Any mention of regulations that directly or indirectly limit access or any other internal or external factors that hinder implementation.
	Overcoming barriers		Any mention of how barriers were overcome, methods supporting the implementation.
	Improvements		Any mention of suggestions, improvements, whether they would recommend it to others.
	Useful applications		Any mention of applications they used most, applications they liked, applications that are easy to understand, and comments on the helpfulness of applications.
	Less useful applications		Any mention of applications they felt like they did not need, disliked.
A-CHESS involvement	Involvement with A-CHESS		Any mention of devices used with A-CHESS, time spent on implementation, A-CHESS use per shift, A-CHESS use off work.
A-CHESS impact	Positive internal impact		Any mention of positive impact of A-CHESS use on client relationship, on workload, daily activities.
	Negative internal impact		Any mention of negative impact of A-CHESS use on client relationship, on workload, daily activities.
	Positive external impact		Any mention of state and/or federal regulations or policy (including Affordable Care Act) that positively impact use of A-CHESS.
	Negative external impact		Any mention of state and/or federal regulations or policy (including Affordable Care Act) that positively impact use of A-CHESS.

Domain	Code	Description
Sustainability	Strategies employed	Any mention of strategies employed for sustainability including strategies to overcome any barriers. Note: When coding, name the strategies employed.
	Monitoring of sustainability	Systems in place: Any mention of processes implemented to routinely measure the impact of A-CHESS on client or organizational outcomes
		Communication: Any mention that A-CHESS impact is shared with staff, leadership
		Frequency: Any mention of the frequency with which A-CHESS results are shared
	Outside environment	Any mention positive or negative of how outside influences, such as regulatory changes or absence of funding to support efforts to bill for technology use, impact sustainability.
	Challenges to sustainability	Any mention of challenges that inhibit sustainability, including staff resistance to A-CHESS. Note: Enter notes about the different barriers or challenges
	Attributes enhancing sustainability	Any mention of A-CHESS attributes enhancing sustainability.
	Attributes inhibiting sustainability	Any mention of A-CHESS attributes inhibiting sustainability. Note: attributes from the client perspective should be also be coded with client population.
	Resources to support sustainability	Any mention of resources (such as staff and funding) needed to sustain changes and how they will be obtained, and how sustainability will be measured. Note: When coding, we should note which strategies were employed.
	Staff engagement	Any mention positive or negative of how staff was engaged in ongoing efforts to sustain A-CHESS.
	Financial impact	Any mention of the business case for sustaining A-CHESS including whether or not the business case (ie, financial gains) associated with A-CHESS have been successful or not successful.
	Facilitator impact	Any mention of which facilitators turned out to be the most important in sustaining A-CHESS within the organization.
Funding	Funding	Any mention of funding or cost, including funding resources to support sustainability, regulations affecting funding or cost, the business case for continuing A-CHESS, expenses, and compensations.
Other technologies	Technology type	Any mention of a new technology that the organization will be or has implemented for their clients.
	Technology platform	Any mention of the platform for which the new technology needs to operate.

Results

Overview

Detailed information about the organizational attributes of the A-CHESS consortium agencies is available in Multimedia Appendix 1. The 14 agencies are geographically distributed equally across the country except for the West region (n=2). The majority of the agencies (9/14, 64%) serve more than 500 clients per year. Table 2 provides an overview of the interview process by agency, including staff interviewed and dates, as well as the date the agency joined the consortium.

A total of 73 interviews were completed with 44 unique individuals. Interviews were conducted with staff (n=24) and leadership (n=20) across the 14 agencies. Table 3 shows that the majority of staff interviewed was female (14/24, 58%); however, 60% (6/10) administrators were male. The staff workforce reported an average tenure of 5.7 years with 46% (6/13) being with their agency less than 3 years. Also, 13 out of the 24 staff (54.2%) reported their age was under 40. Smartphone use was prevalent among the staff members under age 40.

XSL•FO
RenderX

Table 2. Interview timeline by agency.

Agency	Consortium join date	Roles	Initial interview date	12-month interview date	24-month interview date
1	10/1/2011	Admin	6/25/2012	7/23/2013	8/12/2014
		Admin	None	7/10/2013	Several contacts, no response
		Staff	None	8/20/2013	8/25/2014
		Staff	6/26/2012	None	N/A
2	4/1/2011	Admin	2/20/2012	None	N/A
		Admin	4/11/2012	7/23/2013	8/5/2014
		Staff	4/10/2012	7/25/2013	8/4/2014
		Staff	4/10/2012	None	N/A
		Staff	None	7/17/2013	7/30/2014
3	11/1/2011	Admin	None	7/24/2013	7/25/2014
		Staff	6/25/2012	7/24/2013	4 reschedules
		Staff	None	7/16/2013	7/24/2014
		Staff	7/15/2013	None	N/A
4	1/1/2013	Admin	5/1/2014	5/1/2014	
		Admin	6/13/2014	6/13/2014	
5	4/1/2011	Admin	2/29/2012	8/2/2013	Several contacts, no response
		Staff	2/6/2012 (2 staff together)	None	N/A
		Staff		None	N/A
		Staff	None	7/30/2013	8/21/2014
6	11/1/2011	Admin/ Staff	7/5/2012 (admin interview)	7/31/2013 (staff interview)	8/29/2014
		Admin	3/2/2012	None	N/A
		Staff	None	9/6/2013	Not in agency anymore
		Staff	8/7/2012	None	N/A
7	11/1/2011	Admin	6/25/2012	7/30/2013	8/13/2014
		Admin	6/26/2012	7/10/2013	8/12/2014
		Admin	6/26/2012	7/11/2013	not in agency anymore
		Staff	6/21/2012	7/25/2013	Several contacts, no response
		Staff	7/11/2012	None	N/A
8	7/1/2011	Admin	3/5/2014 (24 mos.)	Not in consortium	
		Staff	3/5/2014 (24 mos.)	Not in consortium	
9	4/1/2011	Admin	2/7/2012	Not in consortium	Not in consortium
10	7/1/2011	Admin	6/7/2012	Not in consortium	Not in consortium
11	1/1/2013	Admin	10/28/2013	N/A	
		Staff	11/1/2013	10/23/2014	
12	7/1/2013	Admin	1/15/2014		
		Staff	1/17/2014	1/16/2015	
		Staff	1/28/2014	1/26/2015	
13	11/1/2013	Admin	1/24/2014		
		Admin	2/7/2014	2/25/2015	
		Staff	1/22/2014	1/12/2015	
		Staff	1/27/2014	1/23/2015	

Agency	Consortium join date	Roles	Initial interview date	12-month interview date	24-month interview date
14	5/29/2013	Admin	5/6/2014		
		Staff	5/29/2014	5/29/2014	

Table 3. Qualitative inquiry response rate and respondent demographics.

	Staff	Administrator
Interview time frame, n		
Initial interview	17	19
12-month interview	15	11
24-month interview	5	6
Employee demographics (based on responses received), n (%)		
Gender		
Male	10 (41.7)	12 (60.0)
Female	14 (58.3)	8 (40.0)
Age range		
20-29	5 (20.8)	
30-39	8 (33.3)	
40-49	3 (12.5)	
50-54	2 (8.3)	
Refused/no response	6 (25.0)	
Access/use of technology		
Computers	22 (47.8)	
Mobile phone/smartphone	19 (41.3)	
iPad/iPod	4 (8.7)	
Other (videophone)	1 (2.2)	
Tenure in years		
3 years	6 (46.1)	
4-9 years	5 (38.5)	
9 years	2 (15.4)	
Average tenure in years, mean (SD)	5.7 (5.2)	

Qualitative Analysis

Identification of agencies meeting our definition of short-term sustainability was straightforward. Three agencies (#5, #7, and #13) continued to utilize A-CHESS as an integrated part of their service delivery after the Consortium ended. Three agencies (#8, #9, and #10) dropped out of the Consortium before the 12-month interview. Although the remaining agencies were not able to sustain A-CHESS after the Consortium ended, their efforts were also reviewed to help identify unsuccessful sustainability strategies. We utilized the previously identified sustainability frameworks to organize the qualitative analysis which suggest that sustainability is attributable to the internal organizational supports, the external environment (eg, funding and regulations), and characteristics of the intervention [31-34]. Our analysis focused on (1) the internal organizational supports put in place to sustain A-CHESS, (2) funding as a contextual

influencer from the external environment, and (3) A-CHESS innovation characteristics that support its sustainability.

Internal Organizational Supports

Organizational supports focused on three specific strategies: leadership support, staff engagement, and client engagement. Strong leadership support was present across multiple agencies. Often in these agencies, the leadership team took systematic and organized steps to move forward with the program. One respondent stated, "The whole system had to be developed from the ground up including process development and documentation". In another agency that sustained A-CHESS, leadership understood that using technology to engage clients in treatment is innovative. However, the external environment (eg, reimbursement from insurance providers or Medicaid) is not keeping up with the innovations. As a result, leadership in this agency recognized that external funding to support

A-CHESS would need to come from private donors and/or foundations. The absence of leadership support may hinder development of an A-CHESS business model strategy. A respondent from one agency not sustaining A-CHESS stated that they need to be "able to convince administrators that there is a financial benefit to them".

Staff Engagement

Engaging staff during the initial implementation and ongoing support of an mHealth app like A-CHESS is an important attribute associated with sustainability. The three agencies that were able to sustain A-CHESS used three clearly defined concrete strategies to engage staff (Table 4). One strategy involved the regular review of reports about client use of A-CHESS. This provided staff with additional opportunities to identify clients at risk and to initiate follow-up efforts to re-engage them. One respondent highlighted the importance of monitoring from a clients' perspective when they stated the value of the relapse prevention feature (weekly check-in): "Clients like [A-CHESS], it's a small and valuable intervention". Another staff engagement strategy involved integrating a discussion of A-CHESS into weekly staff meetings. One respondent indicated that "Once a week we sit down and talk and review globally their caseload, [to] extend utilization of A-CHESS". The final staff engagement strategy these three agencies used involved identifying staff who expressed an interest in working with A-CHESS. For example, one respondent indicated that they "selected staff that expressed interest so they're invested so we don't have any staff resistance". A response from one of the agencies highlighted the integration of dedicated staff with the weekly review process when they stated, "We have a team of folks who focuses only on A-CHESS, assess on how clients are using A-CHESS. We have dedicated staff and we monitor client outcome". Other strategies such as providing resources, using peers to provide training, or producing A-CHESS guides were specific to only one of the agencies; while a few of the non-sustaining agencies (#1 and #6) employed one of these key strategies. For example, Agency 1 utilized the A-CHESS reports to provide information to staff about client usage; however, unlike Agencies 5, 7, and 13, that information was not shared with staff on a regular basis. A respondent from one of the agencies that successfully sustained the use of A-CHESS summarized the impact and importance of staff engagement by stating that "we work with patients more at the beginning to sustain interest, help them use the phone…[and]…developed some guidelines". These changes helped patients learn more about how the phone app works. In turn, these individuals were able to help new patients who were not familiar with the phone by providing hands-on support through peer support meetings.

Table 4. Organizational and external environmental strategies supporting sustainability.

	Agency													
	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Organizational level strategies														
Strong leadership support and engagement	√	√		√		√	√					√	√	√
Staff engagement														
Implement reminders related to A-CHESS use for staff	√													
Provide and communicate information through regular review of reports on client usage	√				√		√					√		
Offer training to new staff and ongoing training to all staff including face-to-face meetings to promote staff mentoring opportunities		√	√		√									
Require new staff to be excited and familiar with technology											√			
Produce A-CHESS booklets with screenshots customized to the treatment location or organization specific A-CHESS program guides					√		√							
Establish a web link for A-CHESS demonstrations and video tutorials					√									
Identify staff interest in working with A-CHESS						√	√						√	
Incorporate discussion of A-CHESS into weekly staff meetings						√	√						√	
Involve peers in training							√							
Ongoing client engagement														
Use feedback reports to identify clients not using the system and have staff follow-up	√				√		√						√	
Establish guidelines and contracts related to use of A-CHESS or discharge protocols about how to use A-CHESS	√				√		√							
Use peer counselors or clients to provide training on how to use A-CHESS or share their experiences							√						√	
Adopt A-CHESS materials for specific client needs				√										
Encourage client use of A-CHESS features such as discussion group					√									
Establish task force or internal work group of internal champions including clients to engaged clients in A-CHESS						√	√						√	
Create training sessions for clients including developing user manuals						√	√							
Engage clients early on by showing them how to use the phone							√							
Establish targeted marketing for selected clients	√	√					√				√			
General or generic														
Increase public relations by going to community based gatherings	√													
Used in the agency to augment or replace existing therapeutic approaches					√									
Provide resource supports such as office space and staffing support						√							√	
Work to establish or increase stakeholder awareness of A-CHESS and its benefits						√	√						√	
External environment financial business model														
Leverage billing codes to support reimbursement for A-CHESS													√	
Seek out other sources of funding (eg, donations, mini-grants, etc)			√		√		√						√	
Communicate with insurance carriers or other external agencies to drive support and seek funding													√	
Work with phone carriers to see if lower rates can be secured													√	
Developed a business case for A-CHESS service line						√	√							

Ongoing Client Engagement

Client engagement strategies can promote initial client engagement or ongoing engagement once the client starts to use the technology. While both are important, sustainability relies more on the identification of strategies that encourage ongoing client use of the intervention. Each of the agencies that sustained A-CHESS used two specific strategies to promote client engagement. The first strategy focused on monitoring client engagement with A-CHESS. Staff in these agencies utilized the available reports to identify clients who were not using the system and created processes that required follow-up with the client. As one respondent indicated, "It is up to the staff to engage the people. Peer to peer is not so high so we do need our staff to engage them. Staff participation is very important". In contrast, staff at agencies who did not sustain A-CHESS identified a lack of communication as a barrier to sustainability but did not have a solution to systematize the feedback process or had to determine if these reports were a better source of data than internal tracking systems. As one staff member responded, "The Excel spreadsheet does not give you much unless you know what data to analyze, what variables to use". In some agencies that did not sustain A-CHESS, other organizational changes such as system upgrades or electronic medical record implementation affected staff time to successfully manage the data from A-CHESS.

The second client engagement strategy focused on efforts to establish an agency team (eg, a task force or internal work group) consisting of dedicated staff (and in some instances, clients), whose responsibility was to help promote client engagement with A-CHESS. One agency created a task force of internal champions whose primary role was "all about staying connected to the clients". As a result, staff indicated that A-CHESS helped improve their interactions with clients. A staff member in one of these agencies stated, "It has made my work easier as I can work with patients over the phone and they do not have to come in as often". Another staff member defined ACHESS as "a game-changing event". In this agency, leadership considered "enthusiasm for the program [A-CHESS]" when making counselor hiring decisions.

Other client engagement strategies were used by some but not all of the agencies sustaining A-CHESS. Two of the agencies (#5 and #7) also established protocols, guidelines, or contracts designed to define client expectations for using A-CHESS while in treatment. A respondent from one agency stated that "a client agrees to expectations outlined in a document which he/she signs". The protocols also help clients remain engaged with A-CHESS upon discharge from treatment. Another respondent from a different agency stated, "We offer A-CHESS to our patients in the 28-day inpatient program. It is part of the discharge plan. We talk about it in our weekly alumni group. Before they leave, we give them their phone back (to those who have phones before checking into our inpatient program) and meet with them, put A-CHESS in their phone so they are all set before they walk out of the door". Agency 1 also used this strategy and the feedback reports but did not sustain A-CHESS after the Consortium ended.

In addition, two agencies (#7 and #13) used peer counselors or staff to train clients on A-CHESS use. A respondent from one of these agencies stated, "Peer mentoring staff and clinical staff heavily promote A-CHESS utilization as an important addition to quality and continuum of care". Staff in Agency 7 expanded on these efforts by making sure that clients were trained in how to use A-CHESS. Specifically, staff and peers provided training when the client was first given the phone. While other strategies were applied to promote ongoing client engagement, agencies that sustained A-CHESS did not use them.

External Environment: Financial Business Model

Developing a business model to support the ongoing use of mHealth apps (eg, A-CHESS) is a key strategy for long-term sustainability. Agencies that sustained A-CHESS (#5, #7, and #13) took active steps to reduce costs (eg, provided to clients with a phone) and explored the possibility of additional funding. Leadership in these three agencies used a variety of different strategies to seek or secure additional funding (Table 4). For example, all three agencies sought other sources of funding such as mini-grants or donations. In addition, Agency 13 took advantage of state Medicaid billing codes related to peer monitoring, case management, and crisis intervention to support A-CHESS. This agency also worked closely with the MCOs in the state who "are engaged in A-CHESS promotion and are devoting funds to 'pilot projects' for A-CHESS implementation with their patient population". Two agencies (#5 and 7) that have sustained A-CHESS have taken a different approach to incorporating A-CHESS into their business model. For example, Agency 5 established an A-CHESS service line and developed presentations about A-CHESS impact on client outcomes. Agency leadership uses the presentations to communicate their message to potential donors or foundations, which has helped brand the agency as a leader in the field. They had this to say about their approach: "We've incorporated technology in our service line and we're staying on it. In presentations, we promote and talk about A-CHESS…when I do the presentation and show A-CHESS, I get the audience's attention". As part of the business model strategy for this agency, A-CHESS was offered only to clients with a smartphone and associated data plan, thus avoiding the variable expense associated with buying phones and providing the data service plan.

Absence of a Financial Business Model

Agencies that did not sustain A-CHESS recognized the importance of securing outside funding to support ongoing use of the app. Leadership in three of those agencies (#1, #2, and #11) took active steps to reduce costs and explored the possibility of additional funding including switching carriers, initiating conversations with the state about Medicaid coverage, and using A-CHESS reports to highlight the effectiveness of A-CHESS to the community. Despite these steps, one respondent from these agencies stated, "Funding is always the top barrier. We're always looking for ways to support the A-CHESS program". Another respondent indicated that one of the challenges to sustaining A-CHESS was the "cost of phones; clients aren't really using [A-CHESS] that much the longer they are in treatment". Agencies that relied on grant funding to pay for staff in the A-CHESS pilot were struggling with figuring

out how to continue to pay for those staff in the future. Other agencies that did not sustain A-CHESS did not appear to implement any of the successful business model strategies to help cover costs or secure additional funding. In one of these agencies, the absence of a business strategy combined with financial instability indicated that more work was required "to integrate [A-CHESS] into existing treatment system".

A-CHESS Innovation Characteristics

Attributes of an innovation influence whether it is sustained within an organization. Table 5 summarizes innovation enhancers and inhibitors across all Consortium agencies about specific A-CHESS attributes might contribute to its sustained use their clients.

The top three innovation enhancers mentioned by staff were the discussion group, the ability of A-CHESS to address client needs and staff support, and rapid response to client concerns. Staff believed that A-CHESS allowed the clients to create a sense of community in order to help each other in a time of need. One respondent stated that using A-CHESS "brought all clients to be closer together, they never met each other before but feel connected, they call themselves 'Band of Brothers'". Another respondent said, "It's exciting that technology is part of recovery. It's mobile, on the run, you just press 2 buttons and you can get in touch with people. It allows the client to reach out anytime". Even agencies that were not able to sustain A-CHESS still recognized the importance of engaging clients. Staff in one agency encourages clients to "share in Discussion Group, like when they share in a 12-step meeting…[because it is] Secure, it's safe. It's the recovery version of Facebook. They can communicate and know they are being monitored". A respondent from another agency indicated that A-CHESS "helps them [clients] make appointments, keep up with appointments, [and] continue with their therapy". Despite recognizing how A-CHESS helped with client engagement, these non-sustaining treatment agencies did not actively develop and implement strategies to promote client engagement. The reason identified in one of these agencies was "the failure to put time and work with the clients who didn't use A-CHESS".

Table 5. Attributes of the A-CHESS innovation potentially impacting sustainability.

	Agency													
	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Innovation enhancers														
Discussion group		√	√		√	√	√					√	√	√
Addresses client need (eg, rides, schedules, or information)		√	√		√	√	√					√		√
Staff support and rapid response (eg, BAM survey or panic button)		√	√	√		√	√						√	√
Panic button			√		√	√		√						
Ease of use		√	√	√			√							√
Communication with peers						√						√	√	
Impact on staff workload (eg, makes their jobs easier)		√	√											
Thought of the day			√											√
Weekly survey													√	
My profile		√												
Innovation inhibitors														
Availability and cost of phones		√	√									√	√	
Concerns about being tracked					√								√	
Technological failures (eg, no coverage, features not working etc)		√	√	√		√								
Clients skills or comfort level with smartphones, which may impact regular use of the phone	√					√								

Technology

Technological challenges, phone availability and costs, concerns about being tracked, and the skills of the users were identified as potential innovation inhibitors. During the initial rollout, the phone or features on the phone (eg, meeting locator) did not always work as expected for both clients and staff. One respondent stated, "There are a lot of problems with phones (not good quality). We thought at first it was the clients not taking care of the phones but staff also had problems with phones". Another respondent indicated that "it's too long to load, sometimes get a blank screen… but some clients have complained about not getting signal (connectivity)". These challenges affected clients' use of the phone. For example, one respondent stated, "It is a problem because clients cannot call for transportation". Client comfort level with technology in general could also be a barrier to sustained use. One respondent indicated that "some of the older guys are not tech savvy; we have to teach them over and over again, tell them that they can play with the phone and it won't hurt it". Finally, client concerns about the tracking function of the innovation were raised in a couple of agencies. One respondent stated, "The paranoia effect,

clients think we're tracking them and following them with electronic capability".

Discussion

Principal Findings

Strategies that might influence sustainability may be related to intervention types [37]. A-CHESS is an innovation that requires coordinating multiple staff to implement a new technology in the provider organization. Efforts to successfully sustain such an intervention require support from leadership and a project champion, an understanding of how the innovation fits within the organizational mission, training for all staff involved in the implementation, seeking continued financial resources, integrating a new practice into organization policies and procedures, and continuous monitoring [37]. Our results indicate how these strategies help sustain the use of mHealth apps within the context of addiction treatment programs.

While leadership support is an important part of the different sustainability frameworks [31-34], how it unfolds in the organization is less clear. We found that many agencies indicated the presence of leadership support for A-CHESS implementation. Involved and committed leaders can help facilitate sustainability [38-40]. In agencies that successfully sustained A-CHESS, leadership was actively involved by providing support during implementation. These leaders took steps to ensure that their strategic decisions and plans for A-CHESS created a coherent vision for organizational transformation; empowered employees to participate in the change process, including providing training; and allocated resources, change policies, procedures, and job descriptions [31,41,42]. For example, leadership in these agencies recognized the need to establish A-CHESS use covenants or contracts before providing the innovation to a client.

Fiduciary responsibility is a leadership task. Efforts to secure ongoing funding are an important issue in mHealth implementation [25,43]. Our results indicate that leadership can support efforts for mHealth financing by establishing service lines, leveraging billing codes, and marketing the impact of mHealth on clients to perspective payers. Future research should explore how these different approaches contribute to the long-term sustainability of mHealth apps.

Effectively engaging staff, and in our case, clients, establishes a sense of empowerment that they are a part of the change process, enhances staff motivation to participate in the change process, and leads to more actionable steps by staff to promote sustainment of the change [40,44-48]. We found that specific strategies targeting staff and client engagement differed between agencies that sustained versus did not sustain use of A-CHESS. For example, successful agencies established A-CHESS implementation teams and utilized clients and peer counselors to train new users. As a result, the participation helped to provide sufficient evidence for staff and clients to believe in the benefits of A-CHESS. This local participatory approach enhances sustainability efforts [39,49-51]. Further research is required to understand how staff and client engagement strategies support the sustainment of mHealth apps.

Developing an ongoing measurement and feedback system facilitates and supports organizational sustainability efforts [39,50,52-54]. It has been suggested that such systems have four components: (1) continually and effectively monitor progress of change, (2) keep the organization informed about success, (3) identify new areas for improvement, and (4) establish protocols to communicate and act on the results [55]. Agencies sustaining A-CHESS developed systems focusing on three of the four components, excluding a focus on new areas for improvement. Their strategies integrated monitoring efforts into weekly meetings and established policies and procedures to follow up with unengaged clients. The effectiveness of the approach on client engagement requires further investigation.

Limitations

Providers that participated in the research consortium (a self-selected sample) may not be representative of other treatment providers that might have implemented other mHealth apps. Another limitation related to the sample is that these providers implemented a specific mHealth app—a recovery-oriented app for individuals with substance abuse issues. The strategies associated with implementing and sustaining the A-CHESS app may differ from other mHealth apps. For example, levels of direct care staff engagement associated with use of another mHealth app may require different strategies to promote sustainability.

The qualitative process relied on a sample of interviews from the participating agencies. Staff turnover meant that different staff members were interviewed for the baseline and follow-up surveys. Inconsistency in the timing of the interview is another study limitation. The interviews did not begin when an agency joined the consortium. Time was required to allow the agency to develop a plan to implement A-CHESS in their organization. As such, the initial interview was not conducted until after the plans were developed with a range of 3-16 months between when the organization joined the consortium and when we conducted the initial interview. We then attempted to conduct the follow-up interviews at the appropriate 12-month and 24-month windows. Efforts to schedule the interviews with staff led to, in some instances, the interviews being scheduled as much as 1-5 months outside of the respective 12-month and 24-month timeframe. In addition, the interview process did not include patients using A-CHESS. Their responses about strategies to sustain A-CHESS may have differed from staff. Client perspectives about why they continue to use a specific mHealth technology should be included in future studies.

The external funding strategies are specific to the structure and financing mechanism within the US health care system for substance abuse treatment. The organizational strategies might be transferable to non–US-based substance abuse treatment providers or other types of health care systems. However, the full implementation of the Affordable Care Act in the United States or the actual structure of the payment system in another country may require changes to existing strategies or the identification of new strategies for sustaining an mHealth app in general or A-CHESS in particular within their respective implementation environments. Further research will be necessary to fully understand how different financial strategies and

payment mechanism affect the use of mHealth apps in health care.

Conclusions

As with any technology adoption, the use of mHealth apps for disease management will require a series of changes to the operations in health care provider organizations. Successful organizations will develop adoption plans that address staff and patient engagement, organizational procedures, and business model. Organization leaders will revisit these plans and adapt their implementation process to address emerging needs as they arise. Future research should focus on understanding the environmental, organizational, and innovation attributes that influence an organization's decision to sustain the use of a particular mHealth app for its clients.

Authors' Contributions

All authors were involved in the overall design and implementation of the study. SD and KPR conducted the interviews. EA, JHF, and KAJ coded the interviews. EA, JHF, KAJ, and SD drafted the first version of the manuscript. All authors reviewed and edited the manuscript.

Conflicts of Interest

DHG, KAJ, and SD are shareholders in CHESS Mobile Health, which is a company that markets A-CHESS.

Multimedia Appendix 1

A-CHESS consortium member organizational attributes.

[PDF File (Adobe PDF File), 27KB - jmir_v17i8e201_app1.pdf]

References

1. 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] [Medline: 23349621]
2. Smith A. Smartphone Ownership 2013.: Pew Research Center; 2013 Jun 05. URL: http://www.pewinternet.org/2013/06/05/smartphone-ownership-2013/ [accessed 2014-08-01] [WebCite Cache ID 6TWKmoU4w]
3. Tomlinson M, Rotheram-Borus MJ, Swartz L, Tsai AC. Scaling up mHealth: where is the evidence? PLoS Med 2013;10(2):e1001382 [FREE Full text] [doi: 10.1371/journal.pmed.1001382] [Medline: 23424286]
4. Mikalajunaite E. Mobile Health Market Report 2010-2015: The Impact of Smartphone Applications on the Mobile Health Industry.: research2guidance; 2010 Nov 11. 500m people will be using healthcare mobile applications in 2015 URL: http://research2guidance.com/500m-people-will-be-using-healthcare-mobile-applications-in-2015/ [accessed 2014-08-11] [WebCite Cache ID 6TWKwIBKh]
5. Aitken M, Gauntlett C. IMS Patient Apps for Improved Healthcare: From Novelty to Mainstream.: IMS Institute for Healthcare Informatics; 2013 Oct 01. Report by the IMS Institute for Healthcare Informatics (2013) URL: http://www.imshealth.com/deployedfiles/imshealth/Global/Content/Corporate/IMS%20Health%20Institute/Reports/Patient_Apps/IIHI_Patient_Apps_Report.pdf [accessed 2014-08-11] [WebCite Cache ID 6TWL3Hmg0]
6. McLellan AT. Have we evaluated addiction treatment correctly? Implications from a chronic care perspective. Addiction 2002 Mar;97(3):249-252. [Medline: 11964098]
7. Tai B, Volkow ND. Treatment for substance use disorder: opportunities and challenges under the Affordable Care Act. Soc Work Public Health 2013;28(3-4):165-174. [doi: 10.1080/19371918.2013.758975] [Medline: 23731411]
8. Donker T, Petrie K, Proudfoot J, Clarke J, Birch M, Christensen H. Smartphones for smarter delivery of mental health programs: a systematic review. J Med Internet Res 2013;15(11):e247 [FREE Full text] [doi: 10.2196/jmir.2791] [Medline: 24240579]
9. Dulin PL, Gonzalez VM, King DK, Giroux D, Bacon S. Smartphone-Based, Self-Administered Intervention System for Alcohol Use Disorders: Theory and Empirical Evidence Basis. Alcohol Treat Q 2013;31(3):- [FREE Full text] [doi: 10.1080/07347324.2013.800425] [Medline: 24347811]
10. Gonzales R, Douglas AM, Glik DC. Exploring the feasibility of text messaging to support substance abuse recovery among youth in treatment. Health Educ Res 2014 Feb;29(1):13-22 [FREE Full text] [doi: 10.1093/her/cyt094] [Medline: 24038196]

11. King VL, Brooner RK, Peirce JM, Kolodner K, Kidorf MS. A randomized trial of Web-based videoconferencing for substance abuse counseling. J Subst Abuse Treat 2014 Jan;46(1):36-42 [FREE Full text] [doi: 10.1016/j.jsat.2013.08.009] [Medline: 24035556]

12. Marsch LA, Carroll KM, Kiluk BD. Technology-based interventions for the treatment and recovery management of substance use disorders: a JSAT special issue. J Subst Abuse Treat 2014 Jan;46(1):1-4 [FREE Full text] [doi: 10.1016/j.jsat.2013.08.010] [Medline: 24041749]

13. McTavish FM, Chih M, Shah D, Gustafson DH. How Patients Recovering From Alcoholism Use a Smartphone Intervention. J Dual Diagn 2012;8(4):294-304 [FREE Full text] [doi: 10.1080/15504263.2012.723312] [Medline: 23316127]

14. McKay J. The Use of New Communications Technologies to Evaluate and Intervene in Substance Use Disorders. Curr Behav Neurosci Rep 2014 Jun 20;2(1):23-29. [doi: 10.1007/s40473-014-0017-y]

15. Savic M, Best D, Rodda S, Lubman DI. Exploring the focus and experiences of smartphone applications for addiction recovery. J Addict Dis 2013;32(3):310-319. [doi: 10.1080/10550887.2013.824331] [Medline: 24074196]

16. Walters ST, Ondersma SJ, Ingersoll KS, Rodriguez M, Lerch J, Rossheim ME, et al. MAPIT: development of a web-based intervention targeting substance abuse treatment in the criminal justice system. J Subst Abuse Treat 2014 Jan;46(1):60-65 [FREE Full text] [doi: 10.1016/j.jsat.2013.07.003] [Medline: 23954392]

17. Kuntsche E, Labhart F. Drinking motives moderate the impact of pre-drinking on heavy drinking on a given evening and related adverse consequences--an event-level study. Addiction 2013 Oct;108(10):1747-1755. [doi: 10.1111/add.12253] [Medline: 23692418]

18. Moore SC, Crompton K, van GS, van den Bree M, Bunney J, Lydall E. A feasibility study of short message service text messaging as a surveillance tool for alcohol consumption and vehicle for interventions in university students. BMC Public Health 2013;13:1011 [FREE Full text] [doi: 10.1186/1471-2458-13-1011] [Medline: 24160674]

19. Gustafson DH, McTavish FM, Chih M, Atwood AK, Johnson RA, Boyle MG, et al. A smartphone application to support recovery from alcoholism: a randomized clinical trial. JAMA Psychiatry 2014 May;71(5):566-572 [FREE Full text] [doi: 10.1001/jamapsychiatry.2013.4642] [Medline: 24671165]

20. Glasgow RE, Kessler RS, Ory MG, Roby D, Gorin SS, Krist A. Conducting rapid, relevant research: lessons learned from the My Own Health Report project. Am J Prev Med 2014 Aug;47(2):212-219. [doi: 10.1016/j.amepre.2014.03.007] [Medline: 24953520]

21. Lindquist AM, Johansson PE, Petersson GI, Saveman B, Nilsson GC. The use of the Personal Digital Assistant (PDA) among personnel and students in health care: a review. J Med Internet Res 2008;10(4):e31 [FREE Full text] [doi: 10.2196/jmir.1038] [Medline: 18957381]

22. Kaplan WA. Can the ubiquitous power of mobile phones be used to improve health outcomes in developing countries? Global Health 2006;2:9 [FREE Full text] [doi: 10.1186/1744-8603-2-9] [Medline: 16719925]

23. Mattila E, Orsama A, Ahtinen A, Hopsu L, Leino T, Korhonen I. Personal health technologies in employee health promotion: usage activity, usefulness, and health-related outcomes in a 1-year randomized controlled trial. JMIR Mhealth Uhealth 2013;1(2):e16 [FREE Full text] [doi: 10.2196/mhealth.2557] [Medline: 25098385]

24. Shaw RJ, Bosworth HB, Hess JC, Silva SG, Lipkus IM, Davis LL, et al. Development of a Theoretically Driven mHealth Text Messaging Application for Sustaining Recent Weight Loss. JMIR Mhealth Uhealth 2013;1(1):e5 [FREE Full text] [doi: 10.2196/mhealth.2343] [Medline: 25100678]

25. Abebe NA, Capozza KL, Des Jardins TR, Kulick DA, Rein AL, Schachter AA, et al. Considerations for community-based mHealth initiatives: insights from three Beacon Communities. J Med Internet Res 2013;15(10):e221 [FREE Full text] [doi: 10.2196/jmir.2803] [Medline: 24128406]

26. Steinhubl SR, Muse ED, Topol EJ. Can mobile health technologies transform health care? JAMA 2013 Dec 11;310(22):2395-2396. [doi: 10.1001/jama.2013.281078] [Medline: 24158428]

27. Larimer ME, Palmer RS, Marlatt GA. Relapse prevention. An overview of Marlatt's cognitive-behavioral model. Alcohol Res Health 1999;23(2):151-160. [Medline: 10890810]

28. Witkiewitz K, Marlatt GA. Relapse prevention for alcohol and drug problems: that was Zen, this is Tao. Am Psychol 2004;59(4):224-235. [doi: 10.1037/0003-066X.59.4.224] [Medline: 15149263]

29. Cacciola JS, Alterman AI, Dephilippis D, Drapkin ML, Valadez C, Fala NC, et al. Development and initial evaluation of the Brief Addiction Monitor (BAM). J Subst Abuse Treat 2013 Mar;44(3):256-263 [FREE Full text] [doi: 10.1016/j.jsat.2012.07.013] [Medline: 22898042]

30. Chih M, Patton T, McTavish FM, Isham AJ, Judkins-Fisher CL, Atwood AK, et al. Predictive modeling of addiction lapses in a mobile health application. J Subst Abuse Treat 2014 Jan;46(1):29-35 [FREE Full text] [doi: 10.1016/j.jsat.2013.08.004] [Medline: 24035143]

31. Aarons GA, Hurlburt M, Horwitz SM. Advancing a conceptual model of evidence-based practice implementation in public service sectors. Adm Policy Ment Health 2011 Jan;38(1):4-23 [FREE Full text] [doi: 10.1007/s10488-010-0327-7] [Medline: 21197565]

32. Chambers DA, Glasgow RE, Stange KC. The dynamic sustainability framework: addressing the paradox of sustainment amid ongoing change. Implement Sci 2013;8:117 [FREE Full text] [doi: 10.1186/1748-5908-8-117] [Medline: 24088228]

33.   Schell SF, Luke DA, Schooley MW, Elliott MB, Herbers SH, Mueller NB, et al. Public health program capacity for sustainability: a new framework. Implement Sci 2013;8:15 [FREE Full text] [doi: 10.1186/1748-5908-8-15] [Medline: 23375082]

34.   Scheirer MA, Dearing JW. An agenda for research on the sustainability of public health programs. Am J Public Health 2011 Nov;101(11):2059-2067. [doi: 10.2105/AJPH.2011.300193] [Medline: 21940916]

35.   Damschroder LJ, Aron DC, Keith RE, Kirsh SR, Alexander JA, Lowery JC. Fostering implementation of health services research findings into practice: a consolidated framework for advancing implementation science. Implement Sci 2009;4:50 [FREE Full text] [doi: 10.1186/1748-5908-4-50] [Medline: 19664226]

36.   Bazeley P, Jackson K. Qualitative data analysis with NVivo. Thousand Oaks, CA: Sage Publications Inc; May 29, 2013:1-328.

37.   Scheirer MA. Linking sustainability research to intervention types. Am J Public Health 2013 Apr;103(4):e73-e80. [doi: 10.2105/AJPH.2012.300976] [Medline: 23409904]

38.   Boss RW, Dunford BB, Boss AD, McConkie ML. Sustainable Change in the Public Sector: The Longitudinal Benefits of Organization Development. The Journal of Applied Behavioral Science 2010 Nov 12;46(4):436-472. [doi: 10.1177/0021886310387087]

39.   Bray P, Cummings DM, Wolf M, Massing MW, Reaves J. After the collaborative is over: what sustains quality improvement initiatives in primary care practices? Jt Comm J Qual Patient Saf 2009 Oct;35(10):502-508. [Medline: 19886089]

40.   Savaya R, Elsworth G, Rogers P. Projected sustainability of innovative social programs. Eval Rev 2009 Apr;33(2):189-205. [doi: 10.1177/0193841X08322860] [Medline: 18708528]

41.   Ford JH, Krahn D, Wise M, Oliver KA. Measuring sustainability within the Veterans Administration Mental Health System Redesign initiative. Qual Manag Health Care 2011;20(4):263-279 [FREE Full text] [doi: 10.1097/QMH.0b013e3182314b20] [Medline: 21971024]

42.   Ford JH, Wise M, Krahn D, Oliver KA, Hall C, Sayer N. Family care map: Sustaining family-centered care in Polytrauma Rehabilitation Centers. J Rehabil Res Dev 2014;51(8):1311-1324 [FREE Full text] [doi: 10.1682/JRRD.2014.03.0066] [Medline: 25671632]

43.   Whittaker R. Issues in mHealth: findings from key informant interviews. J Med Internet Res 2012;14(5):e129 [FREE Full text] [doi: 10.2196/jmir.1989] [Medline: 23032424]

44.   Davies B, Tremblay D, Edwards N. Sustaining evidence based practice systemsmeasuring the impacts. In: Bick D, Graham ID, editors. Evaluating the Impact of Implementing Evidence Based Practice. Oxford: Wiley-Blackwell; May 17, 2010:165-188.

45.   Doyle C, Howe C, Woodcock T, Myron R, Phekoo K, McNicholas C, et al. Making change last: applying the NHS institute for innovation and improvement sustainability model to healthcare improvement. Implement Sci 2013;8:127 [FREE Full text] [doi: 10.1186/1748-5908-8-127] [Medline: 24160758]

46.   Johnson K, Hays C, Center H, Daley C. Building capacity and sustainable prevention innovations: a sustainability planning model. Evaluation and Program Planning 2004 May;27(2):135-149. [doi: 10.1016/j.evalprogplan.2004.01.002]

47.   Parker LE, de Pillis E, Altschuler A, Rubenstein LV, Meredith LS. Balancing participation and expertise: a comparison of locally and centrally managed health care quality improvement within primary care practices. Qual Health Res 2007 Nov;17(9):1268-1279. [doi: 10.1177/1049732307307447] [Medline: 17968043]

48.   Scheirer M. Is Sustainability Possible? A Review and Commentary on Empirical Studies of Program Sustainability. American Journal of Evaluation 2005 Sep 01;26(3):320-347. [doi: 10.1177/1098214005278752]

49.   Parand A, Benn J, Burnett S, Pinto A, Vincent C. Strategies for sustaining a quality improvement collaborative and its patient safety gains. Int J Qual Health Care 2012 Aug;24(4):380-390 [FREE Full text] [doi: 10.1093/intqhc/mzs030] [Medline: 22669328]

50.   Parand A, Dopson S, Vincent C. The role of chief executive officers in a quality improvement : a qualitative study. BMJ Open 2013;3(1):- [FREE Full text] [doi: 10.1136/bmjopen-2012-001731] [Medline: 23293245]

51.   Parker LE, Kirchner JE, Bonner LM, Fickel JJ, Ritchie MJ, Simons CE, et al. Creating a quality-improvement dialogue: utilizing knowledge from frontline staff, managers, and experts to foster health care quality improvement. Qual Health Res 2009 Feb;19(2):229-242. [doi: 10.1177/1049732308329481] [Medline: 19092141]

52.   Carlhed R, Bellman C, Bojestig M, Bojö L, Peterson A, Lindahl B, Quality Improvement in Coronary Care (QUICC) Study Group. Quality improvement in coronary care: analysis of sustainability and impact on adjacent clinical measures after a Swedish controlled, multicenter quality improvement collaborative. J Am Heart Assoc 2012 Aug;1(4):e000737 [FREE Full text] [doi: 10.1161/JAHA.112.000737] [Medline: 23130153]

53.   Ovretveit J, Klazinga N. Linking research to practice: the organisation and implementation of The Netherlands health and social care improvement programmes. Health Policy 2013 Feb;109(2):175-186. [doi: 10.1016/j.healthpol.2012.11.005] [Medline: 23270882]

54.   Solberg L. Lessons for non-VA care delivery systems from the U.S. Department of Veterans Affairs Quality Enhancement Research Initiative: QUERI Series. Implement Sci 2009;4:9 [FREE Full text] [doi: 10.1186/1748-5908-4-9] [Medline: 19245709]

55.   Maher L, Gustafson D, Evans A. Sustainability Model and Guide. Leceister, England: British National Health Service Modernization Agency; 2004. URL: http://www.qihub.scot.nhs.uk/media/162236/sustainability_model.pdf [accessed 2015-08-13] [WebCite Cache ID 6akpB7hr6]

Abbreviations

A-CHESS: Alcohol Comprehensive Health Enhancement Support System
BAM: Brief Alcohol Monitor

XSL•FO
RenderX

Original Paper

App Usage Factor: A Simple Metric to Compare the Population Impact of Mobile Medical Apps

Thomas Lorchan Lewis[1], BSc (Hons), MB ChB; Jeremy C Wyatt[2], BA, MBBS, DM, FRCP, FACMI

[1]Kingston Hospital National Health Service (NHS) Trust, London, United Kingdom
[2]Leeds Institute of Health Sciences, University of Leeds, Leeds, United Kingdom

Corresponding Author:
Thomas Lorchan Lewis, BSc (Hons), MB ChB
Kingston Hospital National Health Service (NHS) Trust
Galsworthy Road
Kingston-Upon-Thames
London, KT2 7QB
United Kingdom
Phone: 44 7876453511
Email: tlewis@doctors.org.uk

Abstract

Background: One factor when assessing the quality of mobile apps is quantifying the impact of a given app on a population. There is currently no metric which can be used to compare the population impact of a mobile app across different health care disciplines.

Objective: The objective of this study is to create a novel metric to characterize the impact of a mobile app on a population.

Methods: We developed the simple novel metric, app usage factor (AUF), defined as the logarithm of the product of the number of active users of a mobile app with the median number of daily uses of the app. The behavior of this metric was modeled using simulated modeling in Python, a general-purpose programming language. Three simulations were conducted to explore the temporal and numerical stability of our metric and a simulated app ecosystem model using a simulated dataset of 20,000 apps.

Results: Simulations confirmed the metric was stable between predicted usage limits and remained stable at extremes of these limits. Analysis of a simulated dataset of 20,000 apps calculated an average value for the app usage factor of 4.90 (SD 0.78). A temporal simulation showed that the metric remained stable over time and suitable limits for its use were identified.

Conclusions: A key component when assessing app risk and potential harm is understanding the potential population impact of each mobile app. Our metric has many potential uses for a wide range of stakeholders in the app ecosystem, including users, regulators, developers, and health care professionals. Furthermore, this metric forms part of the overall estimate of risk and potential for harm or benefit posed by a mobile medical app. We identify the merits and limitations of this metric, as well as potential avenues for future validation and research.

Introduction

Overview

The growth in popularity of health and medical apps for health care professionals and patients is widely recognized given their numerous successful uses in a number of health care domains, including clinical health care delivery, education, and health promotion [1-4]. However, a number of concerns regarding the reliability and accuracy of apps have arisen, leading to calls for some form of quality assessment [5-7].

Evaluating the quality of mobile apps is a notoriously difficult problem which currently has no standard solution. Ideally, every medical app should be evaluated and tested by a range of experts to ensure its suitability and applicability to medicine. In practice, this is impossible to achieve given the exponential growth in the app market, low barriers to entry, limited resources, and rapid pace of development [6]. A number of models have been

proposed to help clinicians, app developers, regulators, and commissioning bodies to assess the quality of mobile apps, although it remains to be seen which model has the most utility in practical terms [5,8,9].

A previous paper proposing a framework for risk assessment for mobile medical apps identified many components that increase the potential for harm [5]. The components included inherent factors to the app, such as functionality, content, complexity, and lack of a fail-safe. External factors included the app user, inappropriate usage, inadequate training, and the likelihood of an error being detected. Furthermore, this paper suggested that an important component of the potential harm caused by a medical app is the overall impact that a mobile app has on a given population [5]. Lewis et al noted, "Risk is proportional to the number of patients affected, so disease prevalence or similar indices of the number of people likely to be affected by an error need to be considered."

It follows that a less harmful app used by a large population could pose a greater overall population safety risk than a more harmful app used by a small population. Therefore, it is clear that there is a need to develop a metric that will assess the population impact of mobile medical apps and will allow subsequent comparison across different disciplines.

The Problem: Assessing the Impact of a Specific App

There are a limited number of options currently available to assess the impact of a mobile medical app on any given population; these are shown in Table 1 and are ranked in order of accuracy.

It is important for the various stakeholders to be able to estimate and compare the likely population impacts of specific apps for the reasons shown in Table 2.

It is critical to be able to assess the number of people at risk from an unsafe app at any given time and currently there is no clear method of assessing this. Our objective was to propose, develop, and model a simple metric that can be used to estimate and compare the likely impact of a specific app on a population.

Table 1. Models currently in use for assessing the impact of a mobile app on a given population.

Assessment tool[a]	Examples	Advantages	Disadvantages
Detailed app analytics	High level metrics such as active users, time spent on app, and ethnographic and epidemiological data	Gold standard in terms of data detail Would enable precise population impact to be measured	Large volumes of data Not currently practical Relies on app developers releasing crucial business information Developer bias
mHealth studies [10]	Numerous mHealth studies testing the validity of mobile apps for health care	Rigorous independent trials Often have detailed metrics available Often note the quality of an app	Often focus on one specific app Not many studies available Not easy to compare apps from different disciplines
Number of app downloads	Basic metric available from a number of sources	Can easily compare apps from different disciplines	Information not easily accessible Many users only download an app for trial purposes No information about how often an app is used No information about intended audience Often not reported accurately
Educated guesswork	N/A[b]	Minimal knowledge required to provide estimate	Not accurate or precise Wide observer bias

[a]Assessment tools are ranked in order of accuracy.

[b]Not applicable (N/A).

Table 2. Key reasons for use of population impacts of mobile apps by stakeholders.

Stakeholder	Reason for estimating app impact on population
Regulator	To estimate and compare the overall risks posed if the app is unsafe, and to decide on the appropriate regulatory measures
	To understand the potential for population benefit from effective apps
Guideline developer (eg, NICE[a])	To help understand the impact of an app from a public health perspective
App developer	To justify investment decisions
	To guide update strategy
App users	May use the population impact as a surrogate indicator for quality
Clinicians advising users about the app	May use the population impact as a surrogate indicator for quality
Health insurers and funding schemes	To understand the likely payback from approving reimbursement of the cost of the app
Health economists	As part of an estimate of cost effectiveness of the app
	Could utilize AUF[b] as part of their ranking algorithm
App stores	Surrogate marker for quality

[a]National Institute for Health and Care Excellence (NICE).

[b]App usage factor (AUF).

Methods

Identification of a Simple Metric: App Usage Factor

A broad literature search for existing metric systems in use was carried out. The two authors of this paper (TLL, JCW) searched for relevant papers with regard to their suitability for use when applied to medical apps, however, none were found. The search was expanded to look for metrics in use in other aspects of technology. We felt that there were some useful analogies outside health care; these included *passenger miles* for comparing CO_2 emissions from alternative transport methods, the *readership* of newspapers and magazines (number of sales × number of readers per paper), and common *website metrics*, which capture the number of unique visitors × time spent per visit.

We also brainstormed the criteria for a useful metric. In our view, a good metric should display the following characteristics:

1. Simple to calculate from readily available information

2. Reflects both the number of users and the frequency of use

3. Generates a single, understandable figure within the range of 0 to 10, despite the hugely varying number of users per app (at least a million-to-one ratio)

4. Relatively stable over time for each app

5. Can be used as a denominator for adverse incident reports

6. Makes intuitive sense to users in the same way as does the Richter scale, for example.

7. Has good interobserver agreement and reliability.

Our proposed metric, the app usage factor (AUF), is defined as the logarithm of the product of the total number of active users of a mobile app (A_U) with the median number of daily uses of the aforementioned app (D_U). The formula for calculating the metric can be seen in equation 1 as follows:

$$AUF = \log_{10}(A_U \times D_U) \quad (1)$$

There are a number of points worth considering that enable this metric to fulfil all the desired characteristics. Certain considerations for the measurement of A_U and D_U include geographic boundaries (eg, AUF could be global or country specific), operating system version (eg, AUF could differ for iPhone and Android platforms), and app version number, which would need to be specified or standardized in order to make a meaningful assessment of the AUF.

There are also specific temporal considerations when calculating A_U, D_u, and AUF, in particular, the following:

1. The figure for each variable should be assessed as "stable," for instance, not changed significantly over the past 30 days. This is because the use of apps by users can change extremely rapidly, for example, due to media hype surrounding the release of a new app.

2. The AUF for a specific app can change over time as the app ecosystem evolves. Specific issues that may be encountered include app or operating system updates, which may significantly affect the functionality of the app.

On the basis of the considerations above, it is reasonable to calculate the AUF for a specific app on a quarterly basis, at minimum, or 30 days following a major app update. Given the well-recognized speed of app updates, this may well require further refinement [11].

Validation of the App Usage Factor Metric Against Proposed Criteria

Overview

Any novel metric requires appropriate validation to ensure its suitability and accuracy for the proposed task. A series of computer simulation models were constructed that allowed specific app usage scenarios to be explored. The objective was to validate the AUF metric against the desirable metric criteria, thus confirming its suitability and applicability for practical use.

XSL•FO

RenderX

Three specific scenarios were modelled using Python [12], a high-level, general-purpose programming language, as follows:

1. Exploring stability of AUF as a function of A_u and D_u, including determination of metric limits

2. Simulated app ecosystem model

3. Temporal stability of AUF.

Exploring Stability of the App Usage Factor as a Function of Au and Du, Including Determination of Metric Limits

A 200×200 linear spaced grid was constructed in Python to simulate the behavior of the metric with calculated values of the AUF based on $0 < A_U < 100{,}000$ and $0 < D_U < 50$. A secondary iterative process was used to explore the limits of the AUF while still returning usable results. Specific positive and negative limits were identified.

Simulated App Ecosystem Model

Figures from a recent study suggest that there are now 20,000 health and medical apps available to download from the major online app stores [7]. In order to validate the new metric, Python was used to construct a simulated dataset with 20,000 "apps," each with a random number of daily "uses" and a random number of active "users." The values for A_U and D_U were randomly distributed using a negative exponential probability distribution; this is a continuous probability distribution which describes the time between events in a Poisson process (ie, a process in which events occur continuously and independently at a constant average rate [13]). This probability function was chosen to reflect the decreasing probability of an app being *both* widely used (A_U) *and* used multiple times per day (D_U). The probability distribution function for an exponential distribution is shown in equation 2.

$$P(x) = \lambda e^{-\lambda x} \quad (2)$$

An iterative process was used to identify values for lambda in order to identify suitable limits for the maximum number of daily activities, and the maximum number of active users [14]. This was determined to be 0.00001 and 0.4 for A_U and D_U, giving maximum numbers of 1 million users and 30 uses per day, respectively.

Temporal Stability of the App Usage Factor

In order to assess the temporal stability of the AUF, the behavior of a single app was modelled as a function of time. The strength of the logarithmic component of the AUF is its ability to act as a damping system to external ecosystem factors (eg, media hype). The aim was to show that the AUF would not change dramatically in response to these factors. In order to test this, a series of external ecosystem factors would be applied to the simulation to observe how the AUF changed. Key criteria for the simulation were as follows:

1. D_U calculated to be a random float value between minimum and maximum values of D_U ($D_{U\ MIN}$ and $D_{U\ MAX}$, respectively) according to a uniform probability distribution each day

2. A_U calculated to increase/decrease by x users each day, where x is a float value determined by a uniform probability distribution between minimum and maximum values of A_U ($A_{U\ MIN}$ and $A_{U\ MAX}$, respectively)

3. AUF calculated according to values of A_U and D_U as described by the metric and modelled as a function of time for a period of 2 years.

Specific functions were applied to the model, which intended to simulate the following external app ecosystem events: initial market launch, positive media publicity, negative media publicity, and app/operating system updates. The AUF for the simulated app was then plotted as a function of time to analyze temporal behavior. The values for initial A_U ($A_{U\ INITIAL}$), $A_{U\ MIN}$, $A_{U\ MAX}$, $D_{U\ MIN}$, and $D_{U\ MAX}$ for each external ecosystem event are shown below in Table 3.

Table 3. Initial data used to model the characteristics of the AUF as a function of time for a single mobile app.

External ecosystem event	Day number	$A_{U MIN}$ [a]	$A_{U MAX}$ [b]	$D_{U MIN}$[c]-$D_{U MAX}$[d]
Initial market launch	1[e](No. of users initially set at 50)	-50	50	10-20
Daily market fluctuation	All days other than those below	-50	50	10-20
Positive media publicity	100-110	50	500	10-20
Negative media publicity	350-360	-500	50	10-20
App version/operating system update	501[f](No. of users reset to 500)	-50	50	10-20
Users upgrade to latest version	500-650	-20	250	10-20

[a]Range for minimum number of active users of a mobile app ($A_{U MIN}$).

[b]Range for maximum number of active users of a mobile app ($A_{U MAX}$).

[c]Range for minimum median number of daily uses of an app ($D_{U MIN}$).

[d]Range for maximum median number of daily uses of an app ($D_{U MAX}$).

[e]Initial number of active users of a mobile app ($A_{U INITIAL}$) on day 1 (initial market launch)=50.

[f]At day 501, the number of active users was reset to 500 to simulate app version/operating system update.

Results

Stability of the App Usage Factor as a Function of AUand DU

The results of our model shown in Figure 1 highlight the value for AUF (contour lines) as a function of A_U and D_U. The metric remains stable (ie, AUF > 0) provided $A_U \times D_U$ is greater than 1. It is reasonable to assume that if $A_U \times D_U$ were less than 1, then the specific app would be rarely used, if at all. In practical terms, this relates to a scenario where an app is used regularly once a month by 50 people or less, which was deemed to be an acceptable minimum standard.

Figure 1. A contour plot illustrating the stability of the app usage factor as a function of Au and Du, including determination of metric limits.
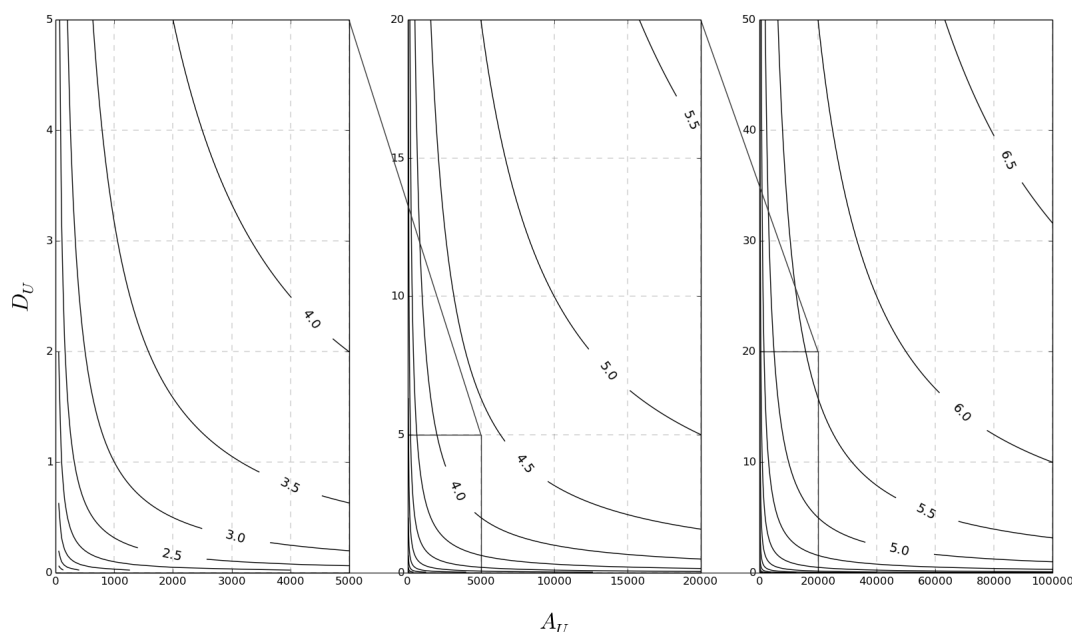
Simulated App Ecosystem Model

The input data can be seen on the left in Figure 2, with the relative frequency of both A_U and D_U plotted as histograms on the right. Each data point represents an individual mobile app with an independent, randomly assigned A_U and D_U. The AUF was then calculated and plotted as a histogram against frequency as shown in Figure 3. As a result of the logarithmic scaling factor, each unit increase in AUF represents a factor of 10 for impact on the population, similar to the Richter scale.

Apps with a similar AUF can be considered to have a comparable population impact to each other, while simultaneously giving a useful indication of the scale of users affected (see Table 4). Distribution of the sample results can be seen in Figure 3. The interquartile range for AUF was calculated to be 4.45 to 5.45. The mean AUF was 4.90, with the standard deviation calculated to be 0.78.

Figure 2. A combined scatterplot (input data, left) and histogram (relative frequency of both Au and Du, right) showing the initial sample dataset of 20,000 mobile medical apps.
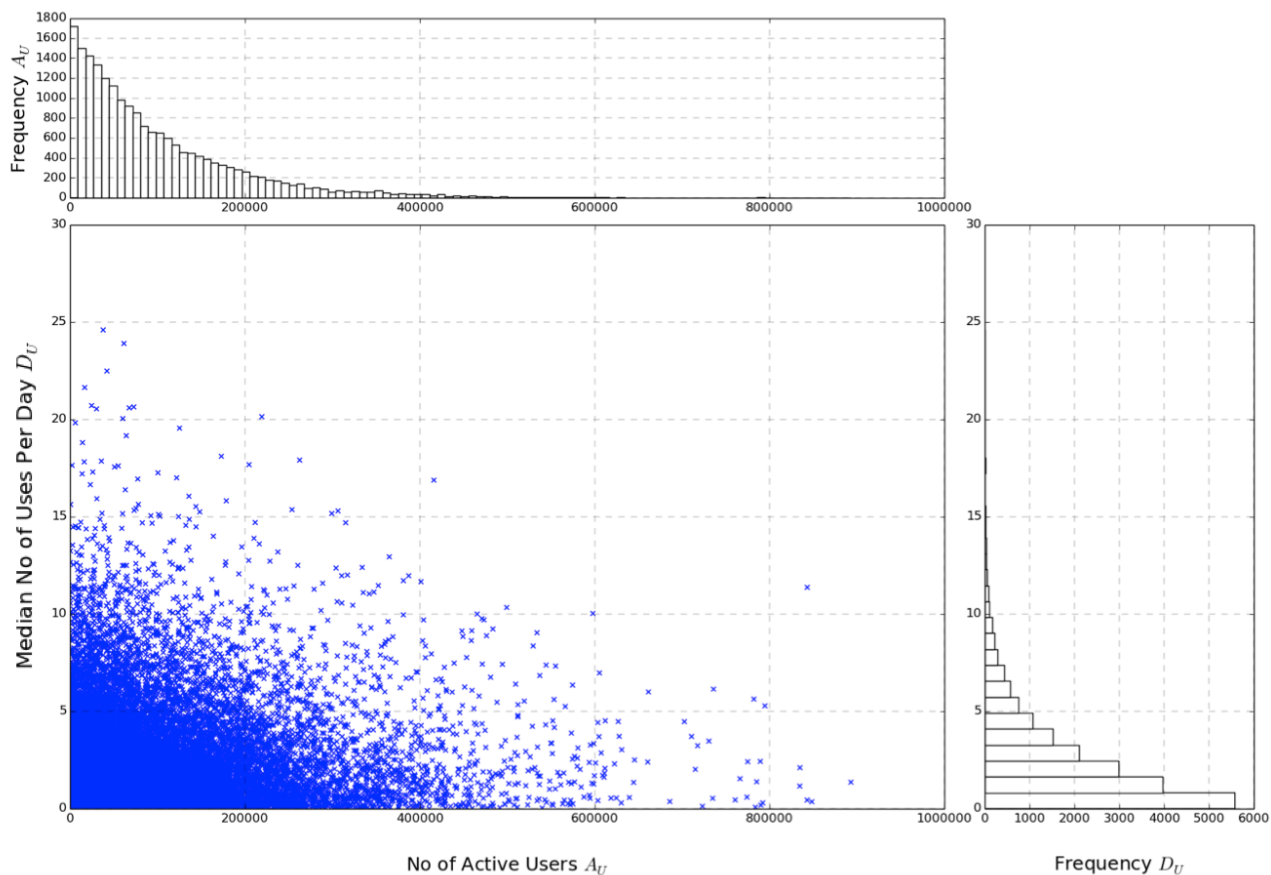
Figure 3. A histogram showing the frequency distribution of the app usage factor for the sample dataset of 20,000 simulated mobile medical apps, including mean and standard deviation for the data.
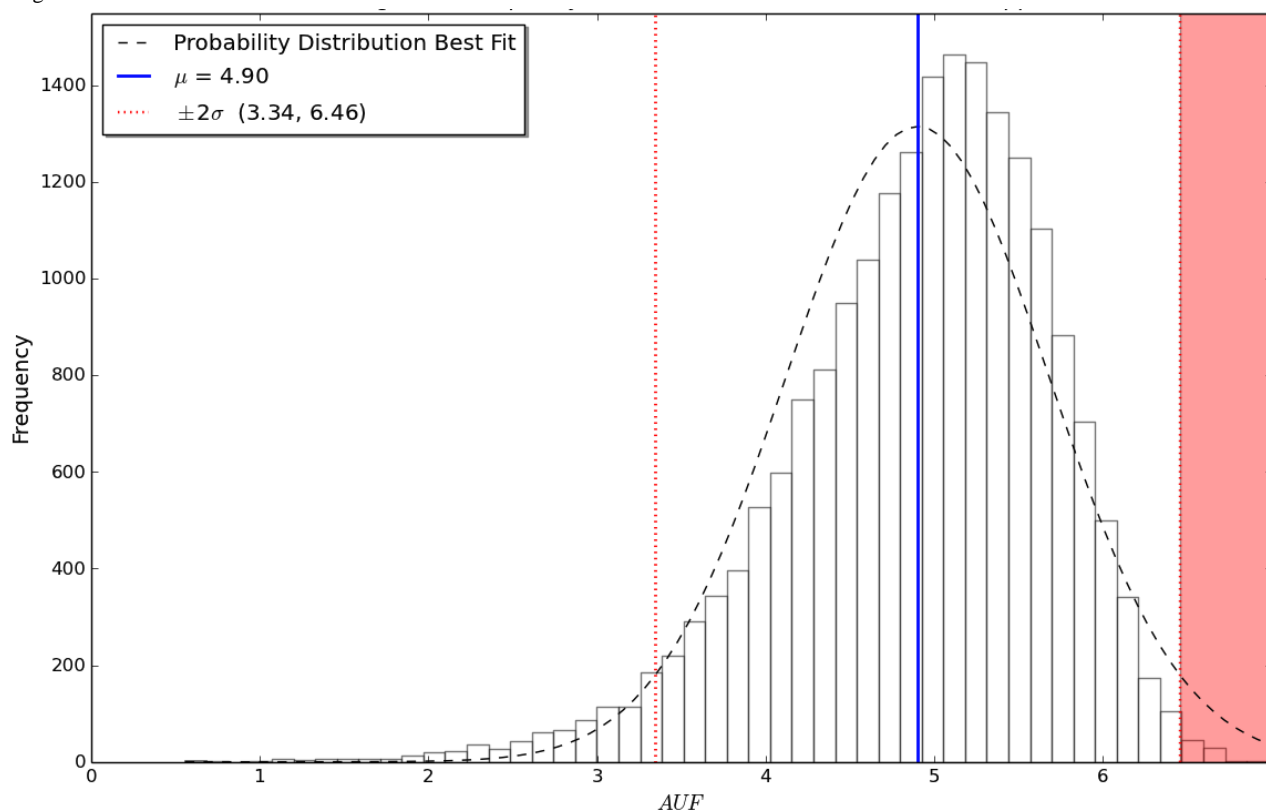
Temporal Stability of the App Usage Factor

The temporal simulation shown in Figure 4 shows that the AUF will tend toward a relatively stable state despite market perturbations. Minor daily fluctuations in A_U and D_U are effectively dampened by the addition of the logarithmic factor. Temporal simulations carried out without the logarithmic factor show an increased sensitivity to small changes in A_U and D_U, which lead to an overall decrease in stability over time. Our simulations suggest that there is a lag time present between a major app ecosystem perturbation and the corresponding change in AUF. This is likely to represent the time taken for information to reach the affected user base and is therefore dependent on the number of users and the magnitude of the market perturbation. It follows that a small market perturbation affecting a small number of users is unlikely to significantly affect the AUF. On the basis of this simulation, it is appropriate to delay measuring AUF for 30 days after any minor market perturbation and 80 to 100 days after a large market perturbation (eg, app launch, operating system update).

Figure 4. A graph showing app usage factor as a function of time for a single mobile app which is subject to a number of simulated app ecosystem events.
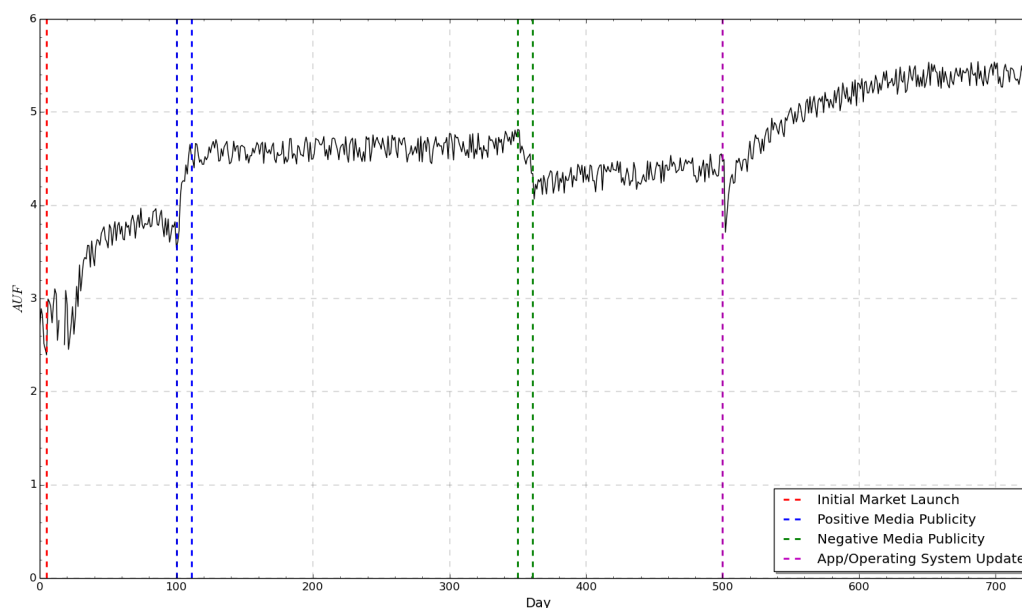
Discussion

Merits and Limitations of the App Usage Factor Metric

Use of the AUF as a metric to assess the population impact of a mobile medical app has many potential benefits for health care professionals, developers, and regulators alike. People who use this metric should be able to make a comparison of the AUF with real-world usage of a particular app, as can be seen in Table 4.

The biggest limitation of this metric is obtaining the two key pieces of information: How many active users there are (A_U) and how many times a day the app is used (D_U). It is not possible to accurately calculate the AUF without approximate figures for these variables. Identification of A_U and D_U could be facilitated through the following means:

1. Voluntary release by app developers/app store companies to an independent body where possible, perhaps as part of a self-certification process. The AUF could be calculated by developers and release of this data does not in itself release confidential information.

2. Compulsory release as part of a formal regulatory process, for example, to gain Food and Drug Administration (FDA) regulation approval.

3. Survey of a target population with subsequent data extraction and extrapolation of log files associated with app usage.

There are a number of further considerations of the app usage factor for risk assessment of mobile apps, particularly when utilizing AUF to estimate population impact of an app.

Risk-based regulatory models such as those utilized by the FDA [15] and the Medicines and Healthcare products Regulatory Agency (MHRA) [16] could potentially target apps with a high AUF (ie, a large population impact) purely on the basis of potential negative impact to a population, as suggested in the red highlighted area in Figure 3. Previous risk assessment analysis of mobile apps identified that approximately 0.5% of apps require this formal regulatory assessment. In our opinion, it is therefore reasonable to identify the apps which have the highest population impact and assess their safety on a case-by-case basis. Identification of these apps is accomplished by identifying all apps that fall more than 2 standard deviations from the mean AUF. In our sample data, this equated to 82 out of the 20,000 simulated apps, a feasible proportion of 0.41% of the apps which could require assessment pending formal regulation; these are highlighted in red in Figure 3. In our simulation, this equates to an app with an AUF greater than 6.46.

A limitation of the AUF in estimating the potential risk of a mobile app is when the number of active users is high within a potentially small user base. This is illustrated by the following thought experiment. A potentially harmful app used regularly once a day by 600 users out of a population of 800 users has a higher chance of causing harm to this user group than the AUF of 2.8 for this scenario alone would suggest. This thought experiment confirms that the AUF is a measure of *population* impact and there are several contextual and other factors that contribute to overall risk posed by a medical app [2].

Table 4. Equivalent population impact of an app based on its corresponding AUF.

App usage factor (AUF)	Equivalent active user daily actions ($AU^a \times DU^b$)
6	1,000,000
5	100,000
4	10,000
3	1000
2	100

[a]Number of active users of a mobile app (AU).

[b]Median number of daily uses of an app (DU).

Conclusion

A key component when assessing app risk and potential harm is understanding the potential population impact of each mobile app. Our new metric would have many potential uses for a wide range of stakeholders in the app ecosystem, including users, regulators, developers, and health care professionals. Furthermore, this metric forms part of the overall estimate of risk and potential for harm or benefit posed by a mobile medical app [2]. We developed and explored the characteristics of a novel but simple, easily calculated metric to assess the likely population impact of a medical app using a sample database of 20,000 apps modelled using a computer simulation. This modelling showed that our proposed metric, AUF, remained stable over time and at extremes of user numbers and daily usage rates, thereby confirming its suitability for further testing in a health care context. We are confident that using this metric will help the population impact of a specific app to be estimated and compared with similar apps. It is important to note that AUF forms but one component of the overall risk and harm potential posed by a specific app. Users should take the AUF into consideration alongside inherent and external risk factors when deciding whether to use an app in clinical practice. For now, the next stage in the validation process is to calculate the app usage factor for a number of health and medical apps using actual usage and population data.

Acknowledgments

Conflicts of Interest

TLL is a writer and editor for iMedicalApps.com, a website dedicated toward providing news on the integration of mobile technology into medical care and the reviewing of medical apps for mobile devices. He does not consult nor receive reimbursement from app developers or creators. JW has no competing interests to declare.

References

1. Lupton D. Apps as artefacts: towards a critical perspective on mobile health and medical apps. Societies 2014 Oct 29;4(4):606-622. [doi: 10.3390/soc4040606]

2. Free C, Phillips G, Watson L, Galli L, Felix L, Edwards P, et al. The effectiveness of mobile-health technologies to improve health care service delivery processes: a systematic review and meta-analysis. PLoS Med 2013 Jan;10(1):e1001363 [FREE Full text] [doi: 10.1371/journal.pmed.1001363] [Medline: 23458994]

3. Ozdalga E, Ozdalga A, Ahuja N. The smartphone in medicine: a review of current and potential use among physicians and students. J Med Internet Res 2012;14(5):e128 [FREE Full text] [doi: 10.2196/jmir.1994] [Medline: 23017375]

4. Lewis TL, Vohra RS. Smartphones make smarter surgeons. Br J Surg 2014 Mar;101(4):296-297. [doi: 10.1002/bjs.9328] [Medline: 24496760]

5. Lewis TL, Wyatt JC. mHealth and mobile medical apps: a framework to assess risk and promote safer use. J Med Internet Res 2014;16(9):e210 [FREE Full text] [doi: 10.2196/jmir.3133] [Medline: 25223398]

6. Misra S, Lewis TL, Aungst TD. Medical application use and the need for further research and assessment for clinical practice: creation and integration of standards for best practice to alleviate poor application design. JAMA Dermatol 2013 Jun;149(6):661-662. [doi: 10.1001/jamadermatol.2013.606] [Medline: 23783150]

7. Aungst TD, Clauson KA, Misra S, Lewis TL, Husain I. How to identify, assess and utilise mobile medical applications in clinical practice. Int J Clin Pract 2014 Feb;68(2):155-162. [doi: 10.1111/ijcp.12375] [Medline: 24460614]

8. Lewis TL. A systematic self-certification model for mobile medical apps. J Med Internet Res 2013;15(4):e89 [FREE Full text] [doi: 10.2196/jmir.2446] [Medline: 23615332]

XSL•FO

RenderX

9. Albrecht U. Transparency of health-apps for trust and decision making. J Med Internet Res 2013;15(12):e277 [FREE Full text] [doi: 10.2196/jmir.2981] [Medline: 24449711]

10. Frisbee K. The King's Fund. 2014 Oct 02. Assessing the impact of mobile health apps URL: http://www.kingsfund.org.uk/audio-video/kathleen-frisbee-assessing-impact-mobile-health-apps [accessed 2015-08-10] [WebCite Cache ID 6aiA0XyQ7]

11. van Velsen L, Beaujean DJ, van Gemert-Pijnen JE. Why mobile health app overload drives us crazy, and how to restore the sanity. BMC Med Inform Decis Mak 2013;13:23 [FREE Full text] [doi: 10.1186/1472-6947-13-23] [Medline: 23399513]

12. Python.: Python Software Foundation URL: https://www.python.org/ [accessed 2015-08-10] [WebCite Cache ID 6ai9o07V0]

13. Weisstein EW. Wolfram MathWorld. Exponential distribution URL: http://mathworld.wolfram.com/ExponentialDistribution.html [accessed 2015-08-11] [WebCite Cache ID 6ajaU1ItA]

14. Payne KB, Wharrad H, Watts K. Smartphone and medical related App use among medical students and junior doctors in the United Kingdom (UK): a regional survey. BMC Med Inform Decis Mak 2012;12:121 [FREE Full text] [doi: 10.1186/1472-6947-12-121] [Medline: 23110712]

15. US Food and Drug Administration. Mobile Medical Applications: Guidance for Industry and Food and Drug Administration Staff. Silver Spring, MD: US Department of Health and Human Services, Food and Drug Administration; 2015 Feb 09. URL: http://www.fda.gov/downloads/MedicalDevices/DeviceRegulationandGuidance/GuidanceDocuments/UCM263366.pdf [accessed 2014-12-10] [WebCite Cache ID 6ai9gIwNc]

16. Medicines and Healthcare Products Regulatory Agency (MHRA). 2010 May 12. Medical Device Technology Forum on the use of software as a medical device URL: http://webarchive.nationalarchives.gov.uk/20141205150130/http://www.mhra.gov.uk/Howweregulate/NewTechnologiesForums/DevicesNewTechnologyForum/Forums/CON084987 [accessed 2015-08-11] [WebCite Cache ID 6ajaM8kA2]

Abbreviations

AU: number of active users of a mobile app
AUINITIAL: initial AU
AU MAX: maximum value of AU
AU MIN: minimum value of AU
AUF: app usage factor
DU: median number of daily uses of an app
DU MAX: maximum value of DU
DU MIN: minimum value of DU
FDA: Food and Drug Administration
MHRA: Medicines and Healthcare products Regulatory Agency
MRC: Medical Research Council
N/A: not applicable
NICE: National Institute for Health and Care Excellence

XSL•FO
RenderX

Original Paper

A Study of the Demographics of Web-Based Health-Related Social Media Users

Shouq A Sadah[1], MS; Moloud Shahbazi[1], MS; Matthew T Wiley[1,2], MS; Vagelis Hristidis[1], PhD

[1]Department of Computer Science and Engineering, University of California, Riverside, Riverside, CA, United States
[2]SmartDocFinder LLC, Riverside, CA, United States

Corresponding Author:
Shouq A Sadah, MS
Department of Computer Science and Engineering
University of California, Riverside
351 Winston Chung Hall
900 University Ave
Riverside, CA, 92521
United States
Phone: 1 9517565691
Fax: 1 9518274643
Email: ssada003@cs.ucr.edu

Abstract

Background: The rapid spread of Web-based social media in recent years has impacted how patients share health-related information. However, little work has studied the demographics of these users.

Objective: Our aim was to study the demographics of users who participate in health-related Web-based social outlets to identify possible links to health care disparities.

Methods: We analyze and compare three different types of health-related social outlets: (1) general Web-based social networks, Twitter and Google+, (2) drug review websites, and (3) health Web forums. We focus on the following demographic attributes: age, gender, ethnicity, location, and writing level. We build and evaluate domain-specific classifiers to infer missing data where possible. The estimated demographic statistics are compared against various baselines, such as Internet and social networks usage of the population.

Results: We found that (1) drug review websites and health Web forums are dominated by female users, (2) the participants of health-related social outlets are generally older with the exception of the 65+ years bracket, (3) blacks are underrepresented in health-related social networks, (4) users in areas with better access to health care participate more in Web-based health-related social outlets, and (5) the writing level of users in health-related social outlets is significantly lower than the reading level of the population.

Conclusions: We identified interesting and actionable disparities in the participation of various demographic groups to various types of health-related social outlets. These disparities are significantly distinct from the disparities in Internet usage or general social outlets participation.

(J Med Internet Res 2015;17(8):e194)   doi:10.2196/jmir.4308

KEYWORDS

online social media; demographics; health forums; health care disparity; drug reviews

Introduction

Background

Social media have been employed in many industries to engage consumers. The health care industry has moved at a slower pace in incorporating social media because of inherent risks such as patient privacy, but recently this rate has increased to fulfill the consumers' needs [1]. Moreover, some companies use social media to provide their employees with wellness videos in order to cut their health care costs [2].

At the same time, health care disparity is a well-studied problem in which two population groups receive unequal services [3]. This problem has been analyzed across various dimensions relating to social determinants in health, including education

XSL•FO
RenderX

and income, environmental hazards, and health outcomes such as mortality, morbidity, and behavioral risk factors [4]. However, health care disparity has not been studied in terms of social media participation. This is important as Internet access and participation in health communities has the potential to improve health outcomes [5]. Hence, understanding the demographics of social outlets, which is the focus of this paper, may shed light on another facet of health care disparity.

To cover different types of Web-based social outlets, we collected data from three types of sources: (1) general Web-based social networks, namely Google+ and Twitter, (2) drug review websites, and (3) health Web forums. We measure the following demographic attributes: age, gender, ethnicity, location, and writing level. Unfortunately, much of this information is unavailable for some, or all, of the sources. For that, we built and evaluated three classifiers for gender, ethnicity, and writing level. User names were used for the gender and ethnicity classifiers. Writing level for users was calculated using modified reading level formula to ignore very long incomprehensible sentences. To extract the location of a post, we use a geocoding application programming interface (API).

Related Work

Analysis of Health-Related Social Outlets

Many researchers have explored the effectiveness of Web-based social media in changing and improving the communication between providers and patients. According to Kane et al [6], 60 million Americans are using Health 2.0 applications, that is, social networks focused specifically on health care. Further, approximately 40% of Americans find an opinion in social media is more trustworthy if it conflicts with a professional's opinion or diagnosis. Hackworth and Kunz [7] found that 80% of American adults have looked online for health-related topics. Recently, there is increased interest in analyzing the health-related content of social media [1]. Denecke and Nejdl [8] analyzed medical concepts mentioned in medical social media posts from different sources to differentiate between informative and affective posts. They found that patients and nurses tend to share personal experiences, while physicians share health-related information. Lu et al [9] studied the content of three disease-specific health communities and their relationship to five informative topics: symptoms, complications, examination, drugs, and procedures. For example, users with breast cancer are more likely to discuss examination, while users with lung cancer are more likely to discuss symptoms. Wiley et al [10] analyzed the content of Web-based social media related to pharmaceutical drugs across several dimensions, including frequently mentioned diseases, keywords, and

sentiment. While the aforementioned work examined health-related social media and content, none of them studied the demographics of the participating users, which is studied in this work.

Measuring and Estimating Demographics of Users of Social Outlets

Survey-Based Methods

In 2012, a Pew Internet Research study showed that women, age 30-49, are more likely to participate in social media websites, where 75% of users are white [11]. eMarketer found that 68.9% of Hispanics use social media compared to 66.2% of the total population. Further, they showed that Hispanics are more likely to compare products online while shopping and write reviews on products [12]. However, no research has focused on health-related social media.

Classifier-Based Methods

Mislove et al [13] built methods to estimate both gender and ethnicity for Twitter users using the 1000 most popular first names reported by the US Social Security Administration and frequently occurring surnames reported by the 2000 US census. Gender and ethnicity methods used the reported first name and last name respectively. Mandel et al [14] analyzed the tweets related to Hurricane Irene using Mislove's gender classifier. We build on Mislove's work when creating our classifiers. While we also classify gender using first names, we extended these methods to screen names when a first name is not present. A related work for estimating reading levels of the US population [15] was presented to discuss limitations of low literacy patients. We measured the writing level based on this work since we did not encounter any similar work.

Methods

Datasets

Our analysis used data collected from three different types of health-related social outlets: general social networks, drug review websites, and health Web forums (Table 1 [16-23]). Google+ and Twitter were chosen as general social networks based on their popularity and number of users (we do not study Facebook because it offers no public interfaces to access its data). For drug review websites and health Web forums, three websites were selected for each, where we considered their breadth of topics and popularity. Figure 1 shows the overall process of our analysis, and Table 1 shows key statistics of each source including number of users, number of posts, and average sentence length. More information about the sources including start and end date is available in Multimedia Appendix 1.

Table 1. Total number of users, posts, and average sentences length for each source.

Dataset	Users, n	Posts, n	Average sentence length (in words)
TwitterHealth [16]	5,095,849	11,637,888	10.82
Google+Health [17]	86,749	186,666	9.03
Drugs.com [18]	74,461	74,461	13.85
DailyStrength/Treatments [19]	213,524	1,055,603	11.92
WebMD/Drugs [20]	122,040	122,040	13.53
Drugs.com/Answers [21]	201,126	5,948,877	6.59
DailyStrength/Forums [22]	165,045	1,128,629	13.2
WebMD [23]	155,912	320,118	15.37

Table 2 shows which of the sources provide data for each of the five demographic attributes. Two demographic attributes are not present in any source: ethnicity and writing level. We therefore created methods to automatically classify these attributes, along with gender for the sources where unavailable. *No* indicates that the demographic attribute is not provided by the source. *Yes* indicates that the demographic attribute is provided by the source. Each classifier uses a distinct part of the user profile as denoted by the table footnotes. The writing level classifier uses the Flesch-Kincaid measure based on all users' posts [24].

Table 2. List of all used sources with the available attributes.

Dataset	Age	Gender	Ethnicity	Location	Writing level
TwitterHealth	No	Gender classifier[a]	Ethnicity classifier[b]	Yes	Writing level classifier
Google+Health	Yes	Yes		Yes	Writing level classifier
Drugs.com	No	Gender classifier[c]	No	No	Writing level classifier
DailyStrength/Treatments	Yes	Yes	No	Yes	Writing level classifier
WebMD/Drugs	Yes	Yes	No	No	Writing level classifier
Drugs.com/Answers	No	Gender classifier[c]	No	No	Writing level classifier
DailyStrength/Forums	Yes	Yes	No	Yes	Writing level classifier
WebMD	No	Gender classifier[c]	No	No	Writing level classifier

[a]First name.

[b]Last name.

[c]Screen name.

To filter health-related posts from Twitter and Google+, we built a list of 276 representative health-related keywords based on five categories:

1. Drugs: First we obtained a list of the 200 most popular drugs by prescriptions dispensed from RxList.com [25]. We then removed variants of the same drug (eg, different milligram dosages) resulting in 125 unique drug names.
2. Hashtags: We selected 11 popular health-related Twitter hashtags such as #HCSM (Healthcare Communications & Social Media).
3. Disorders: We selected 81 popular disorders such as cancer and Alzheimer.
4. Pharmaceuticals: We selected the 12 largest pharmaceutical companies such as Pfizer.
5. Insurance: We selected 44 of the biggest insurances such as Medicare and Humana. A complete list of used keywords can be found in Multimedia Appendix 1.

We used the Twitter streaming API [26], with these keywords as filters, to obtain the relevant tweets for our TwitterHealth dataset. Our Google+Health dataset was collected via the Google+ API [27], where each health-related keyword was used as a query to find relevant posts. For the drug review websites and health Web forums, we built custom crawlers in Java using the jsoup [28] library for crawling and parsing the hypertext markup language (HTML) content. For each source, we collected the available data, including user information, posts, disorder, or condition under which a discussion appears, keywords, tags, etc. We emphasize that we collected only publicly available data in accordance with each site's terms of use; no private data were collected.

Figure 1. Overview of the data collection and analysis process.
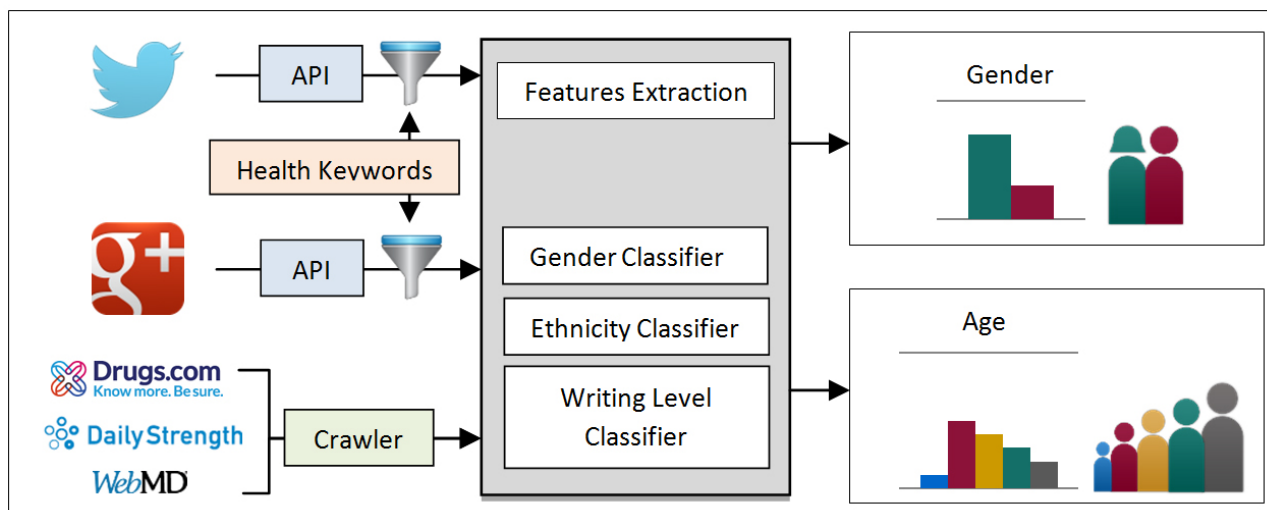
User Demographics Estimation Methods

Overview

We chose five demographic attributes as shown in Table 2: gender, age, ethnicity, location, and writing level. Since these attributes are not available in every source, we created several classifiers to derive missing attributes as specified in Table 2. Note that we do not fill missing values of users for sources that provide this information for at least some of their users, for example, if a user does not provide their age in Google+, we just ignore this user from the age-related analysis. Multimedia Appendix 1 shows the percentages of users who report each attribute in each source.

Gender

Four out of eight sources (Google+Health, DailyStrength/Treatments, WebMD/Drugs, and DailyStrength/Forums) allow users to report their gender (as shown in Table 2). Approximately 80% of the users of these sources chose to report it; thus, the reported gender was used for these sources.

For the other sources where gender is not available, we extended the methods of Mislove et al [13] to classify gender using the reported first name of users, if available; otherwise we extracted first names from user screen names. Note that screen names have not been used before, to the best of our knowledge, for gender estimation. In particular, we first collected the 1000 most popular male and female birth names reported by the US Social Security Administration [29] for each year from 1935 to 1995. Thus, we collected the names of people in 2014 aged 19-79 years old, which constitutes about 73.9% of the population [30]. There are 55,973 unique names in total. We further filtered this list to remove names with an aggregated frequency less than 10,000 or a discriminative gender probability less than 95%. The resultant list contained 1328 names. For TwitterHealth and Google+Health, we checked if one of these 1328 first names is contained in the user-specified name to classify the user's gender. We first cleaned the first name by removing non-alphabetical characters and then performed case-insensitive

string matching. Gender classifier evaluation is reported in Multimedia Appendix 1; the accuracy ranges from 76% to 99%.

Age

Similarly, age was also reported in four sources (Google+Health, DailyStrength/Treatments, WebMD/Drugs, and DailyStrength/Forums).Three sources display the age as a single number, whereas one source displays age as a range (eg, 35-45). Approximately 61% of the users of these sources reported their age. When users provide an age range, the total number of users for each range is distributed uniformly to each year in the range. Ages are then grouped into five age groups: 0-17, 18-34, 35-44, 45-64, and 65 years and older. These age ranges are also used by the US census [31].

Ethnicity

The ethnicity of the users is not reported in any of the sources that we study; therefore, we created an ethnicity classifier similar to Mislove et al [13]. The 2000 US census, which is the most recent available, reports the distribution of ethnicities for each last name (last names with less than 100 individuals were omitted) [32]. For example, the distribution for Hernandez is reported as 4.55% white, 0.38% black, 0.27% Asian, and 93.81% Hispanic. We filtered this list to remove the last names with a frequency less than 1000, or where the discriminative probability of the majority ethnicity is less than 80%. We then use the ethnicity with the majority probability to classify ethnicity based on last name for sources that include the last name of users (Google+Health and TwitterHealth). We understand that race and ethnicity are not the same especially when referring to Hispanics, but in this paper we try to simplify the presentation by only reporting ethnicity, that is, we do not distinguish groups like white Hispanic versus black Hispanic, but only Hispanic. For the other sources (health Web forums and drug review websites), which do not have user names, we found that using the screen name for ethnicity estimation is inaccurate, and hence we do not report on the ethnicity of these sources. Ethnicity labeling and classifier evaluation is reported in Multimedia Appendix 1.

XSL•FO
RenderX

Writing Level

Different methods and formulas for measuring readability are available using different factors such as average number of syllables per words, average number of words per sentences, or average number of letters per words. In our work, we used the Flesch-Kincaid Grade Level [16] formula to estimate the writing level (values generally correspond to school grades 1-12) of the users:

Flesch-Kincaid Reading Age = (0.39 x ASL) + (11.8 x ASW) – 15.59, where ASL is the Average Sentence Length, and ASW is the Average number of Syllables per Word.

Note that since we can only observe the text authored by users, we measure the writing level and not the reading level; however, we use the reading level formula since no alternative formula for the writing level exists. The writing level of a user is computed using the above equation by concatenating all of the user's posts and personal description. Links and hashtags from tweets are removed, and users with less than 100 words in total are ignored. We found that very high reading level was being assigned to users who write very long incomprehensible sentences. This is a case that was not considered by the original Flesch-Kincaid reading age formula, which assumed that the text is grammatically and syntactically correct (eg, the text of a novel). For that, we omit sentences with more than 30 words.

Results

Overview

To put our results in perspective, we compare them with other general demographics statistics. The population and Internet usage for each demographic group was obtained from the US census [29,30], while other statistics for Twitter and Google+ came from other sources [11,33-35]. Further, we compare the demographics of the users participating in health-related discussions on Twitter and Google+ to the overall demographics of the users in these sites. All our results are statistically significant, except the comparison between health Web forums and drug review websites with respect to gender and age group (0-17). Also, there is no significant difference between Google+Health and drug review websites for age group (35-44).

Gender

As shown in Table 3 [11,30,31,33], the gender distribution in the population and Internet usage is almost the same, and there is a slight difference for general social networks. Our first key finding is that drug review websites and health Web forums are dominated by female users; the number of female users is almost four times larger than that of male users. TwitterHealth and Google+Health have similar gender ratios when compared to the overall user base of Twitter and Google+.

The finding that women use health forums much more than men is partially supported by previous research, which shows that women report ill health more frequently than men [36]. In contrast, this is not true for Twitter and Google+, which are dominated by news exchanges [37].

Table 3. Gender distribution for TwitterHealth, Google+Health, drug reviews, health forums, compared to other relevant populations.

Source	Females, %	Males, %
Population [30]	51.05	48.95
Internet Use [31]	51.63	48.37
General social networks [11]	54.68	45.32
Twitter [33]	57.00	43.00
Google+ [33]	37.00	63.00
TwitterHealth[a]	51.81	48.19
Google+Health[a]	35.36	64.64
Drug review websites[a]	78.48	21.52
Health Web forums[a]	78.41	21.59

[a]These results are from this work. Results in the rows above are reported in the respective citations.

Age

Table 4 [30,31,34,38] reports the age distribution of users in the studied social outlets and in other relevant sources, to put the results in perspective. Age groups were chosen based on the US census. We therefore understand that the age ranges are not equal, but since our main goal is comparing the demographics of Web-based health-related social outlets to other statistics such as Internet usage, we chose to follow the census age ranges in computer and Internet access. Further, we provide population distribution in the Table 4 to compare each group size with others. One-fifth of Internet users are in the group 0-17; this percentage drops to approximately 1% for drug review websites and health Web forums. The majority of users on drug review websites are between 45 and 64 years old, and drug reviews have more users over 65 years than any other source; this is expected as older patients use more medications [39]. However, the percentage of drug review users above 65 is slightly lower than the percentage of Internet users over 65, which means that older people still have low participation in Health 2.0 sites.

Also, the 18-34 age group dominates health Web forums, which is congruent with general social networks usage [34]. To summarize, our second key finding is that the participants of health-related social outlets are generally older than those of general-purpose social forums, but still relatively low in the 65+ bracket. This is expected to change in the near future based on the participation statistics in the 45-64 bracket.

Table 4. Age distribution for Google+Health, drug reviews, health forums, and other relevant populations.

Source	0-17 years, %	18-34 years, %	35-44 years, %	45-64 years, %	65+ years, %
Population [30]	24.00	23.11	12.93	26.53	13.44
Internet use [31]	19.30	27.55	14.99	28.36	9.80
General social networks [34]	14.58	27.43	20.68	30.98	6.32
Google+ [38]	8.08	71.61	11.08	7.82	1.42
Google+Health[a]	3.42	53.21	21.89	19.02	2.46
Drug review websites[a]	1.05	31.13	22.36	36.84	8.62
Health Web forums[a]	1.03	39.80	25.81	28.95	4.41

[a]These results are from this work. Results in the rows above are reported in the respective citations.

Ethnicity

For the ethnicity and location analyses, we focus on the US population, in order to compare to available US census statistics. Table 5 [12,31,40,41] shows the results of our ethnicity analysis.

Recall that users' ethnicity in Google+Health and TwitterHealth is classified using our last name-based classifier. Our third key observation is that blacks are underrepresented in health-related social network discussions (Google+Health, TwitterHealth).

Table 5. Ethnicity distribution for TwitterHealth, Google+Health, and other relevant populations.

Source	Asian, %	Black, %	Hispanic, %	White, %
Population [40]	4.5	12.2	15.8	65.1
Internet use [31]	5.5	11.7	13.9	67.2
General social networks [12]	5.3	12.1	14.5	66.5
Twitter [41]	N/A	9	12	71
TwitterHealth[a]	3.24	0.3	23.5	73.0
Google+Health[a]	5.6	0.3	17.4	76.6

[a]These results are from this work. Results in the rows above are reported in the respective citations.

Location

Location is reported in four sources: the two general social networks (TwitterHealth, Google+Health), one drug review website (DailyStrength/Treatments), and one health Web forum (DailyStrength/Forums); approximately 62% of users reported their locations. For TwitterHealth and Google+Health, users report their location using a single string (eg, "NY, NY"). Thus, these strings are further processed to obtain structured locations (eg, state: New York, city: New York). In particular, non-alphanumeric characters and extra spaces were removed, and location strings with a frequency less than 14 were removed. This left us with about 60% of TwitterHealth and Google+Health users with location strings. Each location string was mapped to a location (city, state, country) using the Google Geocoding API [27]. We focus on US users, and hence we remove users from other countries. DailyStrength/Treatments and DailyStrength/Forums list the user's city and state separately; thus, we use the reported state for these sources.

In Figure 2, we show the distribution of users for each type of Web-based health-related social outlet, normalized by state population. Panel A shows the distribution of users in health Web forums, Panel B shows the distribution of users in drug reviews websites, Panel C shows the distribution of users in TwitterHealth, and Panel D shows the distribution of users in Google+Health combined.

To better understand these results, we created Table 6 [31,42-45], which shows the correlation across all states between the normalized (by population) number of users in various health-related social outlets and other societal measures (see Multimedia Appendix 1 for more details). Our fourth key finding is that users in areas with higher income and more access to health care are more likely to participate in Web-based health-related outlets, and particularly in Web forums and drug review sites, which are the primary social sites for health-related information sharing [10]. Further, we see that in Twitter and Google+ the correlation with the number of physicians and education is higher. A reason could be that 59.1% of the 878,194 US active physicians [42] participate in these networks [46], which is a significant number, as the geolocated subsets of the Google+Health and TwitterHealth datasets contain only 882,207 users in the United States. The high correlation with education

may be explained by the high percentage (91%) of Twitter users with college degree or higher [47].

Table 6. Correlation across all states between the normalized (per capita) number of users for each type of health-related social outlets, and each state's population, normalized number of Internet users, normalized number of physicians, normalized number of uninsured patients, average annual income, and percentage of population with college degree or higher.

Correlation	Health Web forums	Drug review websites	TwitterHealth	Google+Health	Google+
Internet usage [31]	0.19	0.28	0.01	-0.01	0.00
No. of physician [42]	0.37	0.19	0.88	0.80	0.44
Uninsured population [43]	-0.40	-0.40	-0.17	-0.11	-0.10
Annual income [44]	0.38	0.27	0.17	0.25	0.26
Education (ratio of people with a college degree) [45]	0.35	0.22	0.56	0.63	0.54

Figure 2. Per state capita number of users in (A) health web forums, (B) drug review websites, (C) TwitterHealth, and (D) Google+Health.
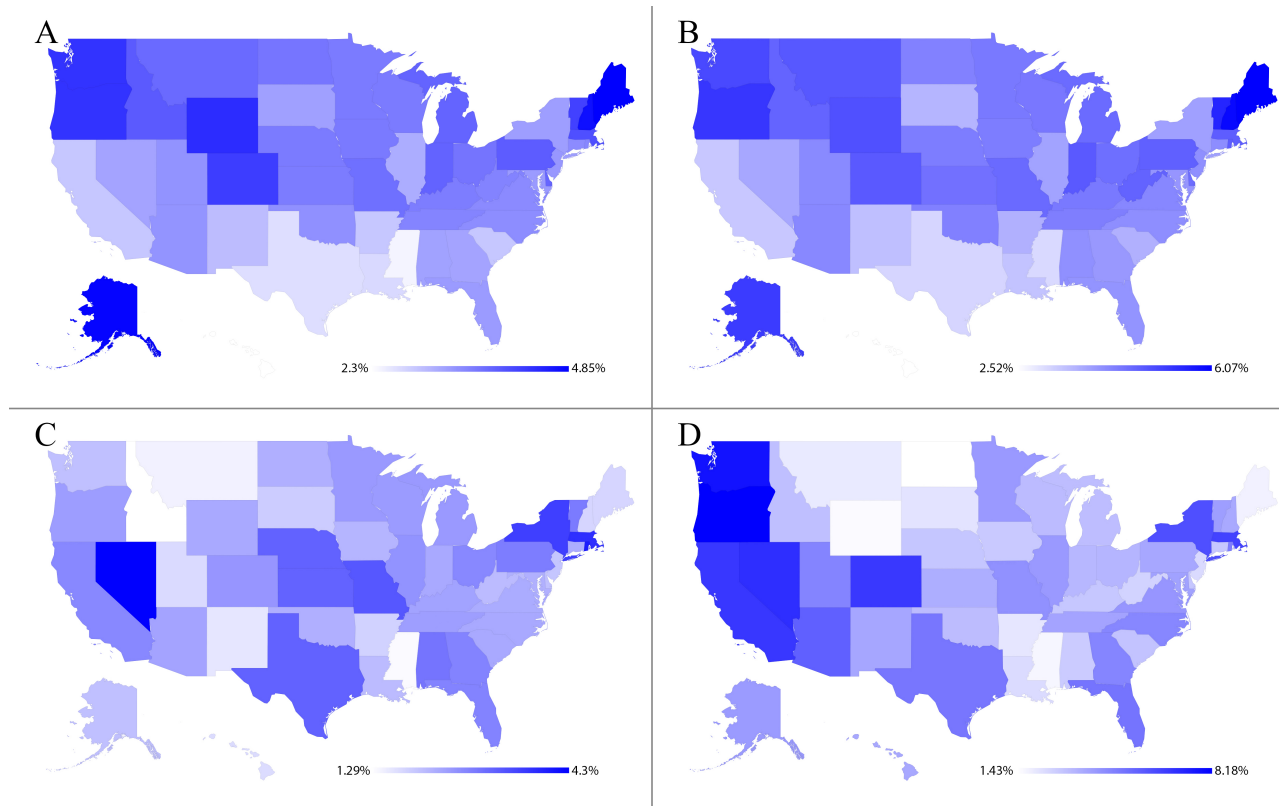
Writing Level

The writing level, as previously mentioned, is measured using a standard reading level formula that assigns a school grade to the given text. For example, when a person writes text at a 5th grade reading level, it implies that their writing should be understood by people that have passed the 5th grade. Table 7 reports our results for writing level of health-related social outlet users. We see that Google+Health users have generally higher writing level than the rest sources, which may mean that more of the Google+Health users are professional accounts.

Next, we try to put these findings in perspective. Unfortunately, related work reports only on reading levels (and not writing levels) of the US population participating in social outlets. Thus, we compare our results in Table 7 to Figure 3, which reports the reading level of the general US population [15].

Table 7. Writing level distribution for TwitterHealth, Google+Health, drug reviews, and health forums.

Source	Age 0-5, %	Age 6-9, %	Age 10-16, %
TwitterHealth	37.77	51.09	11.13
Google+Health	6.45	55.63	37.91
Drug review websites	30.42	66.17	3.41
Health Web forums	28.79	68.24	2.98

Our fifth key finding is that the writing level in health-related social outlets (Table 7) is generally lower than the reading level of the population (Figure 3). Thus users/patients can easily comprehend the posts and hence benefit from the experiences of other users. The benefit of social interaction with respect to health empowerment has been demonstrated before [48]. In an online epilepsy community, 59% of patients found another patient experiencing the same symptoms, 58% had a better understanding of seizures, and 55% learned more about treatments and symptoms.

Figure 3. Reading level of US population.
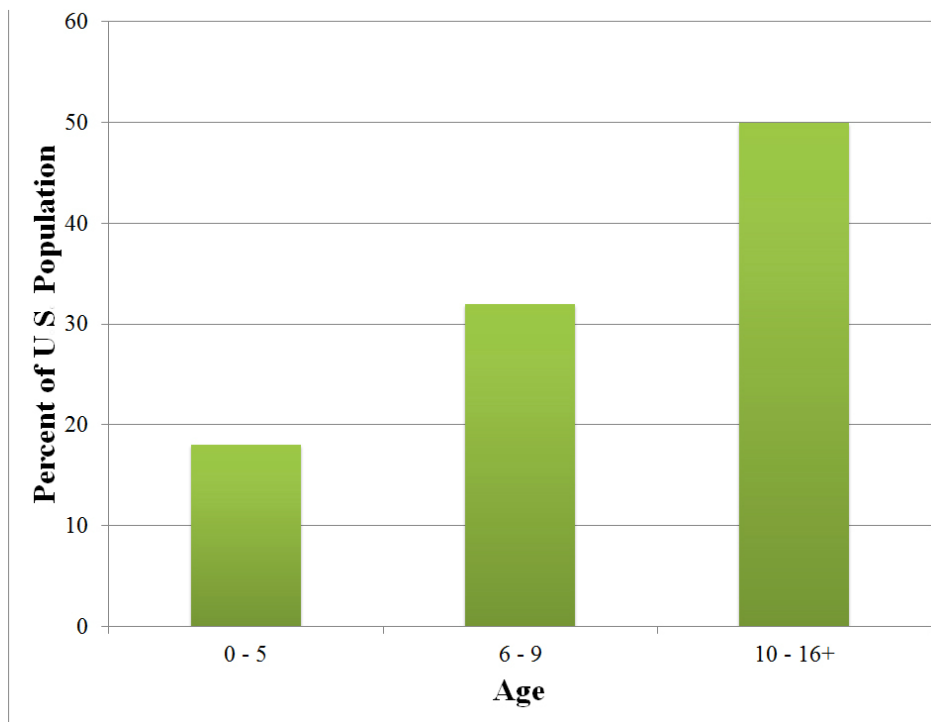
Statistical Significance Tests

Tables 8 and 9 report the *P* values for Pearson's chi-square test of independence and the Mann-Whitney U test. Note that we compute only significance values between sources that we have analyzed and not between our sources and sources analyzed by other works (such as Google+ [38]), since we do not have the raw data for those sources.

Table 8. *P* values for Pearson's chi-square test of independence.

	Gender	Age	Ethnicity	Writing level
TwitterHealth vs Google+Health	<.001	N/A	<.001	<.001
TwitterHealth vs Health Web forums	<.001	N/A	<.001	<.001
TwitterHealth vs Drug review websites	<.001	N/A	<.001	<.001
Google+Health vs Health Web forums	<.001	<.001	<.001	<.001
Google+Health vs Drug review websites	<.001	<.001	<.001	<.001
Health Web forums vs Drug review websites	<.001	<.001	<.001	<.001

Table 9. *P* values for Mann-Whitney U test.

	TwitterHealth vs Google+Health	TwitterHealth vs Health Web forums	TwitterHealth vs Drug review websites	Google+Health vs Health Web forums	Google+Health vs Drug review websites	Health Web forums vs Drug review websites
Gender						
Male	<.001	<.001	<.001	<.001	<.001	.5797
Female	<.001	<.001	<.001	<.001	<.001	.5797
Age						
0-17	N/A	N/A	N/A	<.001	<.001	.5144
18-34	N/A	N/A	N/A	<.001	<.001	<.001
35-44	N/A	N/A	N/A	.01661	.7747	<.001
45-64	N/A	N/A	N/A	<.001	<.001	<.001
≥65	N/A	N/A	N/A	.01066	<.001	<.001
Ethnicity						
White	<.001	<.001	<.001	<.001	<.001	.1316
Black	.6339	<.001	<.001	<.001	<.001	.0944
Asian	<.001	<.001	<.01	<.001	<.001	.8054
Hispanic	<.001	<.001	<.001	<.001	<.001	.6503
Writing level						
0-5	<.001	<.001	<.001	<.001	<.001	<.001
6-9	<.001	<.001	<.001	<.001	<.001	<.001
10-16	<.001	<.001	<.001	<.001	<.001	.00516

Discussion

Principal Findings

Our results can help health care providers customize educational campaigns for different groups. For example, white women should be informed to a larger extent on the possible misinformation spreading in health Web forums, since they participate much more.

Regarding mitigating ethnicity-based health care disparities, we found that Twitter and Google+ are more effective in reaching out to Hispanics about health care offerings. However, this is not true for black ethnicity, who are not overrepresented in any health-related social outlet. This means that there is no single outlet to reach black population, which has been shown to receive worse health care by about 40% comparing to white population [49].

Advertisers may use our results to decide on the best sites to advertise their products; for instance, drug review websites are more appropriate than Google+ to advertise drugs for the 45-64 age bracket, but the opposite is true for the 18-34 age bracket. Further, drug review websites and health Web forums are better to target females when advertising for their products than other health-related social outlets.

In the age results section, we found that younger groups (18-34 years old) participate in large numbers in health forums, which may sound counterintuitive. By analyzing posts for this age bracket, we found the most popular keywords are related to

pregnancy such as birth control, ovulation, and miscarriage. On the other hand, their participation is lower for drug review websites. A possible explanation may be that often patients who talk about pregnancy are not taking any drugs, compared to other conditions like diabetes, where drugs are more common.

We also attempt to explain the disparities in the participation in health-related social outlets based on socioeconomic factors through the state-level participation distributions. Our results in Table 6 show that less access to physicians does not lead to higher participation in health-related social outlets as one would expect. In contrast, it seems that the participation in such outlets is correlated with the access to health care and the average income.

The weak but positive correlation between income and participation to health Web forums and drug review sites may be partially attributed to the higher Internet usage of the more affluent groups, as shown in Table 6. Another possible explanation is that lower income or uninsured persons are more likely to be part of a community with health care disparities [50].The positive correlation between education and participation in health-related social outlets, especially Google+Health and TwitterHealth, may be partially explained by the fact that people with college degrees are less likely to be uninsured, since 10% of college graduates are uninsured, compared to 40% of adults who have not graduated from high school [51]. In addition, 60% of uninsured people are from families with low incomes [51], and the group of people with income lower than US 30K is the lowest group in terms of

accessing health information [52], Hence, our results show that people with low income have less access to health information.

On the other hand, we found that the content in health-related social outlets is easy to understand for almost all users, given the low writing level. That is, the well-known health literacy issue, which is more severe in low-income and lower education populations [5], does not seem to apply to Web-based health-related social outlets. Of course, the low writing level does not address the issue of language, as many low income and low education users in the United States do not speak English at home [53].

Limitations

Our ethnicity and gender classifiers are not perfect, as shown in Multimedia Appendix 1, and thus introduce an error into our analyses. This issue is less significant for gender, since out of all users included in our gender analysis for health Web forums and drug review websites, a majority of the users (over 94%) report their gender, and hence the classifier was only used for 6% of users. Further, a majority of users in drug review websites and health Web forums are female, and our gender classifier obtained an accuracy greater than 99% for females when using a screen name.

Another limitation is the informal writing style of social media posts, as our writing level method uses the average sentence length, which expects that posts are properly punctuated. We addressed this limitation to some degree by only considering sentences of a reasonable length (less than 30 words). Estimating writing level could have been improved by considering other features like typos or spelling mistakes. Further, it would be useful to measure the quality of the posted information, in addition to just the writing level. This is a challenging issue, which we leave as future work.

Since all the attributes are reported by users, there is inevitably self-selection bias. In particular, gender, age, and location are not mandatory in any site. For instance, older people may choose not to report their age. Moreover, choosing to report the real names or posting profile pictures could also create self-selection bias in our gender and ethnicity classifiers. There may also be various types or degrees of bias across different outlets. For instance, WebMD users may use their real name less frequently than Twitter users. This in turn may bias the study results, especially for ethnicity where we depend completely on the classifier results.

Conclusion

We studied user demographics in Web-based health-related social outlets, which we split into three different types: social networks, drug review websites, and health Web forums. The distributions of the demographic attributes—gender, age, ethnicity, location, and writing level—have been analyzed for each source type and compared with relevant baseline user distributions like Internet and general social outlets participation. The results reveal interesting and often unexpected disparities with respect to all demographic attributes.

Authors' Contributions

All authors contributed substantially to this work. They designed and performed the analysis and approved the final version of this manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Online social outlets summary, health keywords, classifiers evaluation, and data coverage.

[PDF File (Adobe PDF File), 36KB - jmir_v17i8e194_app1.pdf]

References

1. ECRI. 2011. Social media in healthcare URL: http://kcmsociety.org/wp-content/uploads/2013/04/ECRI-social-media.pdf [accessed 2015-02-02] [WebCite Cache ID 6W1rk2Yt2]
2. Giving benefits the YouTube treatment - workplace benefits association. 2008. URL: http://www.workplacebenefits.org/news/giving-benefits-youtube-treatment-711171-1.html [accessed 2015-01-31] [WebCite Cache ID 6Vyr2D7tA]
3. Health Disparities. URL: http://www.medicaid.gov/Medicaid-CHIP-Program-Information/By-Topics/Data-and-Systems/Health-Disparities.html [WebCite Cache ID 6W1rr2Maq]
4. Frieden TR, Centers for Disease Control and Prevention (CDC). Forward: CDC Health Disparities and Inequalities Report - United States, 2011. MMWR Surveill Summ 2011 Jan 14;60 Suppl:1-2 [FREE Full text] [Medline: 21430612]

XSL•FO

RenderX

5.   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;16 Suppl 1:77-89. [doi: 10.1080/10810730.2011.596916] [Medline: 21843097]

6.   Kane GC, Fichman RG, Gallaugher J, Glaser J. Community relations 2.0. Harv Bus Rev 2009 Nov;87(11):45-50, 132. [Medline: 19891388]

7.   Hackworth B, Kunz M. Health care and social media: building relationships via social networks. Acad Heal Care J 2010;6(1):55-69 [FREE Full text]

8.   Denecke K, Nejdl W. How valuable is medical social media data? Content analysis of the medical web. Information Sciences 2009 May 30;179(12):1870-1880. [doi: 10.1016/j.ins.2009.01.025]

9.   Lu Y, Zhang P, Liu J, Li J, Deng S. Health-related hot topic detection in online communities using text clustering. PLoS One 2013;8(2):e56221 [FREE Full text] [doi: 10.1371/journal.pone.0056221] [Medline: 23457530]

10. Wiley MT, Jin C, Hristidis V, Esterling KM. Pharmaceutical drugs chatter on Online Social Networks. J Biomed Inform 2014 Jun;49:245-254 [FREE Full text] [doi: 10.1016/j.jbi.2014.03.006] [Medline: 24637141]

11. Pew Research Center's Internet & American Life Project. Social Networking Fact Sheet URL: http://www.pewinternet.org/fact-sheets/social-networking-fact-sheet/ [WebCite Cache ID 6W22L3O2w]

12. eMarketer. 2013. Health care and social media: bulding relationships via social networks URL: http://www.emarketer.com/Article/Hispanic-Shoppers-Bring-Social-Mobile-Habits-Grocery-Aisle/1009839 [accessed 2015-02-02] [WebCite Cache ID 6W21kTHSr]

13. Mislove A, Lehmann S, Ahn Y. AAAI. 2011. Understanding the demographics of Twitter users URL: http://www.ccs.neu.edu/home/amislove/publications/Twitter-ICWSM.pdf [accessed 2015-06-14] [WebCite Cache ID 6ZHoYQIfV]

14. Mandel B, Culotta A, Boulahanis J. A demographic analysis of online sentiment during hurricane irene. Language in Social Media 2012:27-36 [FREE Full text]

15. Harvard. The literacy problem. URL: http://cdn1.sph.harvard.edu/wp-content/uploads/sites/135/2012/09/doakchap1-4.pdf [accessed 2015-06-14] [WebCite Cache ID 6ZHoudAjI]

16. Twitter, Inc. URL: https://twitter.com/ [accessed 2015-08-01] [WebCite Cache ID 6aTENJsel]

17. Google+. URL: https://accounts.google.com/ServiceLogin?service=oz&passive=1209600&continue=https://plus.google.com/?gpsrc%3Dgplp0 [accessed 2015-08-01] [WebCite Cache ID 6aTEQOvFn]

18. Drugs.com. Prescription Drug Information, Interactions & Side Effects URL: http://www.drugs.com/ [accessed 2015-02-02] [WebCite Cache ID 6W23lHwlt]

19. DailyStrength. Treatments: reviews of drugs, therapies and remedies by everyday people URL: http://www.dailystrength.org/treatments [accessed 2015-02-02] [WebCite Cache ID 6W23nXoJa]

20. WebMD Drugs & Treatments - Medical Information and user ratings on prescription drugs and over-the-counter (OTC) medications. URL: http://www.webmd.com/drugs/index-drugs.aspx [accessed 2015-02-02] [WebCite Cache ID 6W23rKUfB]

21. Drugs.com. Medical Questions Answered URL: http://www.drugs.com/answers/ [accessed 2015-02-02] [WebCite Cache ID 6W23vDAi5]

22. DailyStrength. Online Support Groups URL: http://www.dailystrength.org/support-groups [accessed 2015-02-02] [WebCite Cache ID 6W23ztpRe]

23. WebMD - Better information. Better health URL: http://www.webmd.com/ [accessed 2015-02-02] [WebCite Cache ID 6W247u0HQ]

24. The Flesch grade level readability formula. URL: http://www.readabilityformulas.com/flesch-grade-level-readability-formula.php [accessed 2015-02-02] [WebCite Cache ID 6W22mEB3h]

25. RxList - The Internet Drug Index for prescription drugs, medications and pill. URL: http://www.rxlist.com/script/main/hp.asp [accessed 2015-02-02] [WebCite Cache ID 6W24Layfl]

26. The Streaming APIs | Twitter Developers. URL: https://dev.twitter.com/streaming/overview [accessed 2015-06-14] [WebCite Cache ID 6ZHpD7gjO]

27. Google+ API - Google+ Platform — Google Developers. URL: https://developers.google.com/+/api/ [accessed 2015-02-02] [WebCite Cache ID 6W24x2YYc]

28. jsoup Java HTML Parser, with best of DOM, CSS, and jquery. URL: http://jsoup.org/ [accessed 2015-02-02] [WebCite Cache ID 6W2528eFu]

29. Top names over the last 100 years. URL: http://www.ssa.gov/oact/babynames/decades/century.html [accessed 2015-02-02] [WebCite Cache ID 6W25cpwBx]

30. U.S. Census Bureau Demographic Internet Staff. 2012. Age and sex composition in the United States URL: https://www.census.gov/population/age/data/2012comp.html [accessed 2015-02-02] [WebCite Cache ID 6W25iCZh0]

31. U.S. Census Bureau Demographic Internet Staff. 2012. Computer and internet access in the United States URL: https://www.census.gov/hhes/computer/publications/2012.html [accessed 2015-02-02] [WebCite Cache ID 6W271RtOB]

32. US Census Bureau Data Integration Division. Genealogy Data - Frequently occurring surnames from Census 2000 - US Census Bureau URL: http://www.census.gov/topics/population/genealogy/data/2000_surnames.html [accessed 2015-02-02] [WebCite Cache ID 6W27P3i6N]

33. Smith C. User demographics of the major social networks. 2012. URL: http://expandedramblings.com/index.php/user-demographics-of-the-major-social-networks-infographic/ [accessed 2015-02-02] [WebCite Cache ID 6W27uwrNE]

XSL•FO
RenderX

34.   April 2011: The Days of Double-Digit Growth in Social Network Users Are Over. URL: http://www.iab.net/research/
       industry_data_and_landscape/1675/1644724 [accessed 2015-02-02] [WebCite Cache ID 6W284mBJz]

35.   Finn G. Behind The Numbers Of Google+'s Monumental Rise To 25 Million Visitors. 2011. URL: http://searchengineland.
       com/behind-the-numbers-of-googles-monumental-rise-to-25-million-unique-visitor-88076 [accessed 2015-02-02] [WebCite
       Cache ID 6W28agd5M]

36.   Ashley J. BBC News - Women more likely to report ill health than men. 2010. URL: http://news.bbc.co.uk/2/hi/health/
       8588686.stm [accessed 2015-02-02] [WebCite Cache ID 6W28nDUGK]

37.   Kwak H, Lee C, Park H, Moon S. What is Twitter, a Social Network or a News Media? ACM 2010 Apr:591-600 [FREE
       Full text]

38.   The comprehensive resource for Google+ trends and statistics. URL: http://www.gplusdata.com/users/sign_in [accessed
       2015-02-02] [WebCite Cache ID 6W29H2CHP]

39.   Gu Q, Dillon CF, Burt VL. Prescription drug use continues to increase: U.S. prescription drug data for 2007-2008. NCHS
       Data Brief 2010 Sep(42):1-8 [FREE Full text] [Medline: 20854747]

40.   US Census Bureau. 2012. Statistical abstract of the United States - Section 1 URL: http://www.census.gov/prod/2011pubs/
       12statab/pop.pdf [accessed 2015-02-01] [WebCite Cache ID 6W29OSGmX]

41.   Bennett S. Twitter, Facebook and LinkedIn: age, ethnicity and gender of the major social networks. 2011. URL: http://www.
       adweek.com/socialtimes/pew-social-network-age-ethnicity-gender/453165?red=at [accessed 2015-02-02] [WebCite Cache
       ID 6W2A8l2Cu]

42.   Young A, Chaudhry H, Thomas JV, Dugan M. A census of actively licensed physicians in the United States. J Med Regul
       2012;99:11-24 [FREE Full text]

43.   Denavas-walt B, Proctor B, Smith J. US Census Bureau, Current Population Reports. Washington, DC; 2013. Income,
       poverty, and health insurance coverage in the United States URL: http://www.census.gov/prod/2013pubs/p60-245.pdf
       [accessed 2015-02-02] [WebCite Cache ID 6W2D1eglk]

44.   US Census Bureau Data Integration Division. Income URL: http://www.census.gov/hhes/www/income/data/statemedian/
       [accessed 2015-02-02] [WebCite Cache ID 6W2D9aO6n]

45.   US Census Bureau. 2012. Statistical abstract of the United States - Section 4 URL: http://www.census.gov/prod/2011pubs/
       12statab/educ.pdf [accessed 2015-02-02] [WebCite Cache ID 6W2DHYzQJ]

46.   Cooper CP, Gelb CA, Rim SH, Hawkins NA, Rodriguez JL, Polonec L. Physicians who use social media and other
       internet-based communication technologies. J Am Med Inform Assoc 2012;19(6):960-964 [FREE Full text] [doi:
       10.1136/amiajnl-2011-000628] [Medline: 22634078]

47.   Hampton KN, Goulet LS, Rainie L, Purcell K. Pew Research Center's Internet & American Life Project. 2011 Jun 16.
       Social networking sites and our lives URL: http://www.pewinternet.org/2011/06/16/social-networking-sites-and-our-lives/
       [accessed 2015-08-01] [WebCite Cache ID 6aTEsGHIL]

48.   Wicks P, Keininger DL, Massagli MP, de la Loge C, Brownstein C, Isojärvi J, et al. Perceived benefits of sharing health
       data between people with epilepsy on an online platform. Epilepsy Behav 2012 Jan;23(1):16-23 [FREE Full text] [doi:
       10.1016/j.yebeh.2011.09.026] [Medline: 22099528]

49.   Agency for Healthcare Research & Quality (AHRQ). 2014. Disparities in healthcare quality among racial and ethnic minority
       groups URL: http://www.ahrq.gov/research/findings/nhqrdr/nhqrdr10/minority.html [accessed 2015-06-14] [WebCite
       Cache ID 6ZHq5iyFV]

50.   Mendes E. Gallup poll. 2010. In US, health disparities across incomes are wide-ranging URL: http://www.gallup.com/poll/
       143696/health-disparities-across-incomes-wide-ranging.aspx [accessed 2015-02-02] [WebCite Cache ID 6W2E9gDMj]

51.   Adler N, Newman K. Socioeconomic Disparities In Health: Pathways And Policies. Health Affairs 2002 Mar 01;21(2):60-76
       [FREE Full text] [doi: 10.1377/hlthaff.21.2.60]

52.   Madden M. Pew Research Center. 2013. Technology use by different income groups URL: http://www.pewinternet.org/
       2013/05/29/technology-use-by-different-income-groups/ [accessed 2015-06-14] [WebCite Cache ID 6ZHs8GsTY]

53.   Ryan C. US Census Bureau. 2013. Language Use in the United States URL: http://www.census.gov/prod/2013pubs/acs-22.
       pdf [accessed 2015-02-01] [WebCite Cache ID 6W2EHGxIj]

Abbreviations

API: application programming interface

XSL•FO
RenderX

Original Paper

The Impact of Social Media on Dissemination and Implementation of Clinical Practice Guidelines: A Longitudinal Observational Study

Pushpa Narayanaswami[1], MBBS, DM; Gary Gronseth[2], MD; Richard Dubinsky[2], MD, MPH; Rebecca Penfold-Murray[3], BA, MPP; Julie Cox[3], BA, MFA; Christopher Bever Jr[4], MD, MBA; Yolanda Martins[5], PhD; Carol Rheaume[3], MSPH; Denise Shouse[3], BA; Thomas SD Getchius[3], BA

[1]Beth Israel Deaconess Medical Center, Department of Neurology, Harvard Medical school, Boston, MA, United States

[2]University of Kansas School of Medicine, Department of Neurology, Kansas City, KS, United States

[3]American Academy of Neurology, Minneapolis, MN, United States

[4]VA Maryland Health Care System's MS Center of Excellence, Department of Neurology, University of Maryland School of Medicine, Baltimore, United States

[5]Dana Farber Cancer Institute, Survey and Data Management Core, Harvard Medical School, Boston, MA, United States

Corresponding Author:
Pushpa Narayanaswami, MBBS, DM
Beth Israel Deaconess Medical Center
Department of Neurology
Harvard Medical school
Neurology TCC-8
330 Brookline Ave
Boston, MA, 02215
United States
Phone: 1 617 667 8130
Fax: 1 617 667 3175
Email: pnarayan@bidmc.harvard.edu

Abstract

Background: Evidence-based clinical practice guidelines (CPGs) are statements that provide recommendations to optimize patient care for a specific clinical problem or question. Merely reading a guideline rarely leads to implementation of recommendations. The American Academy of Neurology (AAN) has a formal process of guideline development and dissemination. The last few years have seen a burgeoning of social media such as Facebook, Twitter, and LinkedIn, and newer methods of dissemination such as podcasts and webinars. The role of these media in guideline dissemination has not been studied. Systematic evaluation of dissemination methods and comparison of the effectiveness of newer methods with traditional methods is not available. It is also not known whether specific dissemination methods may be more effectively targeted to specific audiences.

Objective: Our aim was to (1) develop an innovative dissemination strategy by adding social media-based dissemination methods to traditional methods for the AAN clinical practice guidelines "Complementary and alternative medicine in multiple sclerosis" ("CAM in MS") and (2) evaluate whether the addition of social media outreach improves awareness of the CPG and knowledge of CPG recommendations, and affects implementation of those recommendations.

Methods: Outcomes were measured by four surveys in each of the two target populations: patients and physicians/clinicians ("physicians"). The primary outcome was the difference in participants' intent to discuss use of complementary and alternative medicine (CAM) with their physicians or patients, respectively, after novel dissemination, as compared with that after traditional dissemination. Secondary outcomes were changes in awareness of the CPG, knowledge of CPG content, and behavior regarding CAM use in multiple sclerosis (MS).

Results: Response rates were 25.08% (622/2480) for physicians and 43.5% (348/800) for patients. Awareness of the CPG increased after traditional dissemination (absolute difference, 95% confidence interval: physicians 36%, 95% CI 25-46, and patients 10%, 95% CI 1-11) but did not increase further after novel dissemination (physicians 0%, 95% CI -11 to 11, and patients -4%, 95% CI -6 to 14). Intent to discuss CAM also increased after traditional dissemination but did not change after novel dissemination (traditional: physicians 12%, 95% CI 2-22, and patients 19%, 95% CI 3-33; novel: physicians 11%, 95% CI -1 to

XSL·FO
RenderX

-21, and patients -8%, 95% CI -22 to 8). Knowledge of CPG recommendations and behavior regarding CAM use in MS did not change after either traditional dissemination or novel dissemination.

Conclusions: Social media-based dissemination methods did not confer additional benefit over print-, email-, and Internet-based methods in increasing CPG awareness and changing intent in physicians or patients. Research on audience selection, message formatting, and message delivery is required to utilize Web 2.0 technologies optimally for dissemination.

Introduction

Evidence-based clinical practice guidelines (CPG) are statements that assist clinical decision making by providing recommendations for optimizing patient care for a specific clinical question. They are informed by a systematic review of evidence and an assessment of the benefits and harms of the evaluated care options [1,2]. Dissemination of CPG recommendations is commonly undertaken by publishing the CPG in a peer-reviewed journal, sending email or paper notices to physicians, and advertising through news media outlets [3].

There is a large amount of medical information available on the Internet and social media. There is some evidence that social media are useful for disseminating medical information, but the quality and accuracy of information vary and may even be misleading [4-22]. Specific to neurology, one report noted accurate demonstration of the Epley maneuver for benign paroxysmal positional vertigo on YouTube [23]. Web-based tools and social media networks (eg, Facebook, YouTube, LinkedIn, Twitter, Google+) have the potential to reach large audiences in a short time span for rapid communication of CPG recommendations [5]. In a recent survey by the Pew Research Center, over two-thirds of Americans reported using the Internet for health and fitness information [4]. This figure did not vary substantially across demographic subgroups. However, there is scant research on the effectiveness of the use of social media platforms in dissemination and implementation of CPG recommendations. The use of targeted approaches within social media platforms to direct and focus guideline dissemination efforts to specific populations has not been evaluated [24].

The American Academy of Neurology (AAN) has developed CPGs since 1989 and has employed a formal dissemination program since 1999 to raise awareness and enable implementation of CPG recommendations. In this study, we evaluate the effectiveness of social media for disseminating recommendations of the recently developed CPG, "Complementary and alternative medicine in multiple sclerosis" ("CAM in MS") [25]. Complementary and alternative medicine (CAM) use is widely prevalent in 33-80% of patients with multiple sclerosis (MS). These patients often do not discuss this use with their physicians [26-33]. A study of information sources used by people with MS revealed that the Internet was the first source of general health information in 73% and for MS-specific information in 59% [34]. Because this is the first CPG on CAM use in MS, we used it to study dissemination tactics, with special emphasis on social media use, to inform future dissemination efforts for CPGs.

The specific aims of this study were to (1) develop an innovative dissemination strategy by adding novel, social media–based methods to traditional dissemination methods for the CPG "CAM in MS", (2) evaluate whether the addition of social media–based methods improves CPG awareness and knowledge of CPG recommendations in the two target audiences of patients and physicians/clinicians (referred to herein as "physicians"), and (3) evaluate whether the addition of social media–based methods improves the implementation of CPG recommendations. Implementation in this context is defined as the adoption and integration of evidence-based health interventions to change practice [35].

Methods

Study Definitions and Design

This was a longitudinal observational study, using quantitative survey methods (AHRQ IR18HS022004-01; Grants.gov tracking 11129815). The study was determined to be exempt from the need for ethical review and approval by the Committee on Clinical Investigations, Beth Israel Deaconess Medical Center, Boston. We have substantially complied with the Workgroup for Intervention Development and Evaluation Research (WIDER) recommendations for reporting research evaluating behavioral interventions [36]. The intervention and co-intervention were traditional and novel dissemination methods, respectively, to disseminate the AAN CPG "CAM in MS". We defined traditional methods as dissemination using print-, email-, and Internet-based methods. This included publication of the CPG in *Neurology*, the official journal of the AAN; issuance of a news release (electronic release and public relations pitch to approximately 700 science and medical reporters); development of clinician and patient summaries, a clinical case example, and presentation slide set of CPG content; news articles in AAN*news, Neurology Today*, and *Neurology Now* (official publications of the AAN); and electronic notices to 26,965 AAN members, consisting of all-member emails, an AAN *e-News* announcement, and an announcement in the AAN quarterly Leadership Update and highlights of the CPG on AAN.com.

We defined novel methods as dissemination through social media platforms. These included an audio podcast on *Neurology*; videos for patients and physicians posted on YouTube, with links to the CPG on AAN.com; and Facebook, Twitter,

LinkedIn, and YouTube digital advertising (where feasible, targeted audiences endorsing an interest in CAM use in MS and related terms were selected; Multimedia Appendix 1). For digital advertisements, the research team treated two of the more popular channels as primary (Facebook and Twitter) and two as secondary (LinkedIn and YouTube) for the purpose of allocating funding over the 90-day dissemination period. Over three periods of 30 days each, the team tested digital advertisements across all four platforms, mixing and matching written text (copy) and visual images, and evaluated the number of impressions (the number of times the advertisement was "served", or seen) for each advertisement. The goal was first to identify which copy would produce the most impressions in the first period and which image would product the most impressions in the second period, and then to promote the most impactful copy and image for the final period. We also held a chat on Twitter in partnership with *TIME* Magazine, the National Multiple Sclerosis Society (NMSS), and Beth Israel Deaconess Medical Center in Boston between 12 noon and 1 p.m. Eastern Time on August 28, 2014. A moderator led the Twitter chat by asking questions regarding general information about MS, what treatments are available, what the evidence indicates, and where

future research should be directed. Informational emails were sent during both traditional and novel dissemination periods to patient organizations (Multiple Sclerosis Association of America, Multiple Sclerosis Foundation, NMSS). These emails informed the organizations of the availability of the guideline and patient summaries, provided links to the guideline, gave a summary of the key guideline recommendations, and requested their support in disseminating the guideline to their members.

The study design and timeline are summarized in Figure 1. We assessed outcomes using survey questionnaires developed for this study. We conducted four surveys in each of the two target populations (physicians, patients). First, we conducted a pre-dissemination survey 1 month before dissemination of the CPG (Feb 2014). The CPG was then published in *Neurology* in March 2014 and simultaneously disseminated using traditional methods. Three months after traditional dissemination (June 2014), the guideline was disseminated using novel methods. We conducted the second and third surveys, post-dissemination survey-1 and post-dissemination survey-2, 1 month after traditional (April 2014) and novel (July 2014) dissemination, respectively. Finally, 6 months after traditional dissemination (Sept 2014), we conducted post-dissemination survey-3.

Figure 1. Timeline of dissemination (horizontal axis), CPG awareness, correct CAM knowledge, intent to discuss CAM, and behavior for patients and physicians. Patient results are represented in the upper set of graphs, and physician results in the lower set. The horizontal axis represents time in months and delineates the point of the traditional and novel dissemination efforts and each of the 4 study surveys: pre-dissemination, post-traditional dissemination (post-1), post-novel dissemination (post-2), the results 6 months after traditional dissemination (post-3), and combined results of all post-dissemination surveys. The vertical axis represents the percentage of respondents. The outcome variables (awareness, knowledge, intent, and behavior) are represented by orange, blue, green, and maroon lines, respectively. Dotted lines indicate significant differences between the 2 surveys. The vertical bars represent within-surveys comparison of unaware and aware respondents for each outcome variable. Starred boxes indicate significant differences.
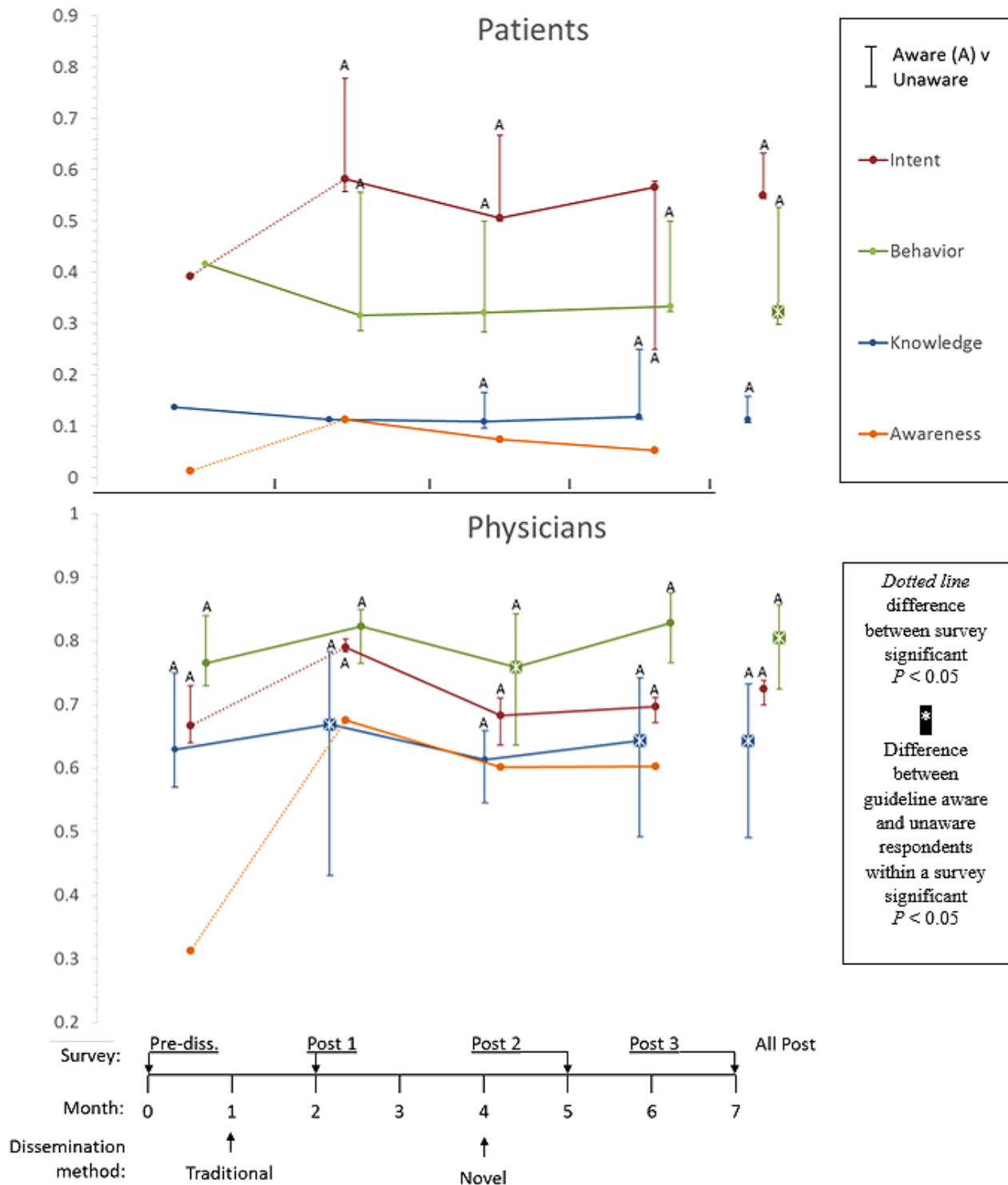
Study Population

The study participants were drawn from two populations: physicians who treated patients with MS, and patients with MS / caregivers of patients with MS. We identified physician participants from the AAN membership database using the following criteria: neurologist or neurology advanced practice provider with a focus in MS practice, residing in the United States, aged ≤68 years, with mailing and email addresses, who had not received an AAN survey in the prior 6 months. A

random sample of 620 physicians was drawn from this population for the pre-dissemination survey. For each subsequent survey, physicians surveyed in the previous survey(s) were excluded, and 620 physicians were randomly selected from the remaining population.

Patient participants were identified from the database of subscribers to *Neurology Now*, an official bimonthly publication of the AAN for patients. Of 23,568 subscribers who self-identified as interested in MS, 10,818 subscribers with email addresses were used as the final dataset. From this dataset, 200 patients were randomly selected for the pre-dissemination survey. For each subsequent survey, patients surveyed in the previous survey(s) were excluded, and 200 patients were randomly selected from the remaining population.

Surveys and Data Collection

A team of investigators and survey methodologists developed the surveys in accordance with the American Association of Public Opinion Research evidence-based, best practices for survey research [37] (see Multimedia Appendix 2 for survey development methods and survey data collection procedure). For patients, the surveys evaluated their knowledge and use of CAM in general, the specific CAM therapies discussed in the CPG, reasons for CAM use, and perceived efficacy of CAM. We explored potential confounders: frequency of Internet use and the presence of any disability interfering with Internet use. In the physician surveys, we assessed knowledge, attitudes, and behavior regarding CAM in MS.

Data collection was performed by the Dana-Farber/Harvard Cancer Center Survey and Data Management Core, who used DatStat Illume. Survey responses were assigned a numeric code and were stored in a separate database without identifying information.

Statistical Analysis

Outcome Measures

The outcome measures were a change in knowledge, intent, and behavior, all with respect to CAM in MS. Knowledge of CAM was assessed by six survey questions regarding the effectiveness of magnetic therapy, oral cannabis extract, smoked cannabis, ginkgo biloba, hyperbaric oxygen, and bee venom therapy, all of which were discussed in the CPG. A 5-point ordinal scale was used for responses (Multimedia Appendix 3). For analysis of change in knowledge, we used the response to the question: "Taking ginkgo biloba orally is effective for improving memory in people with MS". Change in intent was defined as patients' intent to discuss CAM use with their physicians and physicians' intent to discuss CAM use with their patients. Change in behavior was defined as patients' discussing CAM with their physicians and physicians' recommending that their patients start or stop any CAM therapy. General attitudes (beliefs and views towards CAM in MS) were also explored.

The ordinal responses were collapsed into two categories: (1) agree and somewhat agree, and (2) somewhat disagree, disagree, and not sure. We believed that this question measured knowledge most accurately because responses to it were unlikely to be biased by prior opinions as compared with responses likely to be elicited by the therapies assessed in the other questions.

The primary outcome was the difference in respondents' intent to discuss CAM for the two study populations (ie, physicians and patients) with their respective patients/physicians, after novel dissemination (post-dissemination survey-2) as compared with that after traditional dissemination (post-dissemination survey-1). Secondary outcomes were (1) awareness of the CPG, (2) knowledge of CPG recommendations, (3) behavior regarding CAM use in MS (these three outcomes were measured at baseline, ie, pre-publication/dissemination, and after the two dissemination periods, traditional and novel, ie, between-surveys differences), and (4) knowledge, intent, and behavior, measured in CPG-aware physicians and patients. All four outcomes were compared with those unaware of the CPG, within each of the four surveys (ie, within-survey differences). Because behavior changes take time and may not be captured in the time frame of this study and because knowledge may not translate into action, we selected intent to change a priori as a surrogate outcome of behavior. The underlying framework for this choice is the Theory of Planned Behavior that is used to predict behavioral intention and behavior. Intention leads to behavior when attitudes are strong and perceived behavioral control is high [38].

Sample Size Estimations

Physicians

Assuming a 10% responder rate per survey and a 20% change in intent as clinically important, we surveyed 620 physicians in each of four surveys to obtain 62 responses, for an 80% chance (β) of detecting this difference at a significance level alpha=.05.

Patients

Assuming a 20% responder rate per survey and a 10% change in intent as clinically important, we surveyed 200 patients in each of four surveys to obtain 19 responses, for an 80% chance (β) of detecting this difference at a significance level alpha=.05.

Data Analysis

As the measure of effect, we used the absolute difference (AD) between surveys in the proportion of respondents who (1) were aware of the CPG, (2) had correct knowledge of the recommendations discussed in the CPG, (3) endorsed intent to discuss CAM use in MS, and (4) endorsed a recent behavior change regarding CAM use. Secondary analysis stratified respondents within each survey into CPG-aware and CPG-unaware groups. We compared intent, knowledge, and behavior of respondents in CPG-aware and CPG-unaware groups (within survey differences). Precision was measured with 95% confidence intervals (CI). We did not adjust for multiple comparisons for the secondary outcomes.

Results

Dissemination Efforts

The results of the traditional and novel dissemination efforts are summarized in Multimedia Appendix 4.

Results of Surveys

We surveyed a total of 2480 physicians (620 per survey) and 800 patients (200 per survey) across four surveys. The *total response rate* across all four surveys for physicians was 25.08% (622/2480), and for patients was 43.5% (348/800), over twice our predicted rate, and was stable across all four surveys (physicians 23-27%, patients 42-45%) (Table 1). Approximately equal numbers of participants responded to mailed (paper) surveys and email surveys (mailed survey responses: physicians 52-60%, mean across surveys 55%; patients 47-54%, mean across surveys 50.5%). Response rate denominators to individual questions may vary slightly because of incomplete responses. There was no substantial difference between respondent and nonrespondent physicians across the surveys in age or sex (Table 1). Comparison of patient respondents to nonrespondents was not possible, as demographic data were not available for patients from the *Neurology Now* database.

Table 1. Characteristics of surveyed patients and physician (95% confidence interval rounded to nearest whole percent).

	Physicians, % (95% CI)	Patients, % (95% CI)
Sampled, n	2480	800
Responded, n	622	348
Age		
Age sample years, mean (SD)	52.0 (9.4)	
Age respondents years, mean (SD)	52.1 (9.8)[a]	54.8 (11.7)
Women		
Sampled	31.4 (30-33)	
Respondents	34.4 (31-38)[b]	80.2 (76-84)
White	74.4 (71-78)	86.8 (83-90)
Internet use <daily	4.7 (3.3-6.6)	25.6 (21-30)
Practice >15 years	61.9 (58-66)	
Care for >50 patients with MS	57.4 (54-61)	
University-based practice	21.1 (18-24)	
Ever recommend any CAM	79.6 (76-83)	
MS duration >10 years		56.3 (51- 61)
College diploma		55.4 (50- 61)
Walking without assistance		56.1 (51-61)
No difficulty with hand use		43.9 (38-50)
Normal vision		49.2 (44-55)
Ever used any CAM		70.9 (66-76)

[a]Difference in mean age between sample and respondents with 95% CI -0.1 years, -0.93 to 0.73.

[b]Difference in percent women between sample and respondents with 95% CI 3%, -1.1 to 7.2.

Demographics and General Attitudes to Complementary and Alternative Medicine in MS

Physician Respondents

The mean age (standard deviation) was 52.1 (SD 9.8) years (range across surveys 50.7; SD 10.3 to 52.2 years; SD 9.7). Over a third (199/578, 34.4%; range across surveys 31-38%) were women, and three-quarters (445/598, 74.4%; range 69-82%) were white. The practice focus was 31.7% (185/583) group practice, 21.4% (123/583) university-based practice, 18.7% (109/583) solo practice, and 17.0% (99/583) multispecialty practice, and the remaining were in health maintenance organization (HMO), government, or other practice types. A total of 62.1% (362/583) of physicians had been in practice for >15 years, and 12.5% (73/585) for <5 years. Over half the physicians (337/587, 57.4%) were treating more than 50 patients with MS, and almost all (572/600, 95.3%) used the Internet daily (Table 1).

Approximately half the physicians reported routinely discussing CAM with their patients (307/598, 51.3%, range across surveys 48-56%). Very few (23/607, 3.8%; range 1-6%) reported being unaware of CAM use in their patients, and 79.6% (483/607) of physicians said they recommended some form of CAM to their patients (range 76-83%). Yoga (total 385/607, 63.4%, mean across surveys 61%), massage therapy (total 417/607, 68.7%, mean 52%), and acupuncture (total 408/607, 67.2%, mean 51%) were the CAM therapies that physicians considered most useful. Physicians were also most likely to recommend these three CAM therapies (mean across surveys massage 49%, yoga 48%, acupuncture 47%). Half (313/607, 51.6%) of the physicians reported that their patients used marijuana, but only 6.9% (42/607) stated that they would recommend it.

XSL•FO
RenderX

Patient Respondents

The mean age was 54.8 (SD 11.7): range across surveys 54.2 (SD 12.2) to 55.6 (SD 11.6) years. A total of 80.2% (243/303; range 75-87% across surveys) were women. Three-quarters (236/317, 74.4%, range 68-83%) used the Internet daily, and only 9.5% (30/317; range 8-11%) used the Internet once a week or less. Some disability in use of the Internet was reported by just under a quarter (68/312, 21.8%; range 18-28%). Over half the patients reported MS duration of >10 years (171/304, 56.3%; range across surveys 53-60%). MS duration was <5 years in 16.8% (51/304; range 14-19%), 55% (range 48-62%) had at least a college diploma (college diploma: 82/221, 37.1%; some post-graduate education; 28/275, 10.2%, post-graduate diploma: 58/245, 23.7%), over half (169/301, 56.1%; range 49-67%) were ambulatory without assistance, half (148/301, 49.2%; range 37-56%) had normal vision, and under half (133/303, 43.9%; range 34-47%) had no difficulty with use of their hands.

Across surveys, 67.7% (212/313) of patient respondents had heard of CAM (55/79, 70% in the pre-dissemination survey; 49/77, 64%, 46/80, 58%, and 62/77, 81% in the three post-dissemination surveys, respectively) (Table 1). The most common therapies patients reported discussing with their physicians were acupuncture (26/233. 11.2%), dental amalgam removal (21/233, 9.0%), and bee sting therapy (16/233, 6.9%).

Comparison of Results of Pre-dissemination, Post-Traditional Dissemination, and Post-Novel Dissemination Surveys

Awareness of the clinical practice guideline "Complementary and Alternative Medicine in Multiple Sclerosis"

In the pre-dissemination survey, 31.3% (45/144) of physicians were aware of a guideline on CAM in MS. In physicians, there was a significant increase in awareness of the CPG after traditional dissemination (pre-dissemination survey vs post-dissemination-1 survey, awareness AD 36%, 95% CI 25-46%). Although the heightened awareness of the CPG persisted after novel dissemination, there was no further increase in awareness after novel dissemination (AD post-dissemination -1 vs post-dissemination-2 surveys -7%, 95% CI -18 to 4) (Figure 1).

Only 0.1% (1/76) patients was aware of the upcoming CPG in the pre-dissemination survey. There was a statistically significant increase in awareness of the CPG after traditional dissemination that did not change after novel dissemination, although the absolute number of aware patients was small (AD pre-dissemination vs post-dissemination-1: 10%, 95% CI 1-11; post-dissemination-1 vs post-dissemination-2: -4%, 95% CI -6 to 14).

Intent to Discuss Complementary and Alternative Medicine

The intent to discuss CAM (for physicians, with their patients with MS, and for patients/caregivers, with their physicians) increased significantly in both groups after traditional dissemination (AD pre-dissemination vs post-dissemination-1: physicians were 12%, 95% CI 2-22, patients were 19%, 95%

CI 3-33). Our primary outcome measure did not change: the proportion of either patients or physicians reporting an intent to discuss CAM did not increase after novel dissemination as compared with that after traditional dissemination (AD post-dissemination-1 vs post-dissemination-2: physicians were -11%, 95% CI -1 to -21, patients were -8%, 95% CI -22 to 8).

Stratified analysis of the differences in intent between CPG-aware and CPG-unaware respondents within each survey revealed that awareness of the CPG was not associated with an increase in physicians' reporting of an intent to discuss CAM with their patients (intent change, physicians, AD aware vs unaware, all surveys: 4%, 95% CI -5 to 13). CPG-aware patients in post-dissemination surveys were also not significantly more likely to report an intent to discuss CAM with their physicians (intent change, patients, AD aware vs unaware, all surveys: 9%, 95% CI -14 to 28). However, due to the small number of aware patients, the precision of the estimated difference is low.

Knowledge of Clinical Practice Guideline Recommendations

Correct knowledge of the recommendations discussed in the CPG did not significantly change in either physicians or patients after traditional or novel dissemination as compared with pre-dissemination (physicians AD pre-dissemination vs post-dissemination-1: 4%, 95% CI -7 to 15; post-dissemination-1 vs post-dissemination-2: -6%, 95% CI -16 to 5; patients AD pre-dissemination vs post-dissemination-1: -2%, 95% CI -13 to 8; post-dissemination-1 vs post-dissemination-2: -4%, 95% CI -14 to 6).

In the stratified analysis, physicians who were aware of the CPG had better knowledge of CAM than physicians who were not aware of the CPG (knowledge change, physicians, AD aware vs unaware, all surveys: 24%, 95% CI 15-33). Patients who were aware of the CPG did not have significantly better knowledge than those who were not aware (knowledge change, patients, AD aware vs unaware, all post-dissemination surveys: 5%, 95% CI -6 to 27). However, because of the small number of aware patients, the precision of this estimate is low.

Behavior With Regard to Complementary and Alternative Medicine Therapies

The number of physicians who had recommended that their patients stop or start using any CAM therapy did not change significantly after either traditional or novel dissemination as compared with pre-dissemination; that is, there was no change in behavior (behavior change AD pre-dissemination vs post-dissemination-1: 5%, 95% CI -4 to 14; post-dissemination-1 vs post-dissemination-2: -6.4%, 95% CI -16 to 3). Across all surveys, physicians who were aware of the CPG made recommendations more frequently regarding the use of CAM therapies as compared with CPG-unaware physicians (AD physicians making recommendations, aware vs unaware, all surveys: 13%, 95% CI 6-21).

There was no significant increase in the proportion of patients who reported having recently discussed CAM with their physicians following traditional or novel dissemination (behavior change, patients, pre-dissemination vs post-dissemination-1, AD: 10%, 95% CI -6 to 26;

post-dissemination-1 vs post-dissemination-2: 1%, 95% CI -14 to 15). However, across all surveys, CPG-aware patients were more likely to have discussed CAM with their physicians (AD patients, aware vs unaware, all surveys: 23%, 95% CI 1-44).

Discussion

Diffusion and Dissemination

The term *diffusion* is used by some authors to describe the distribution and unaided adoption of information, whereas *dissemination* refers to a more active process of communication to improve knowledge [39]. Other authors use the terms synonymously. Dissemination is defined as "the purposive distribution of information and intervention materials to a specific public health or clinical practice audience." Dissemination research studies how "information about health promotion and care interventions is created, packaged, transmitted, and interpreted among important stakeholder groups" [40]. It is recognized that there is a gap between CPG development and the delivery of care in practice. Merely reading a CPG rarely leads to implementation of recommendations [39]. Active, effective dissemination of CPG recommendations to end users is essential for optimizing care delivery.

Principal Results

In this study, about a third of the physicians were aware of the "CAM in MS" CPG before publication and dissemination, and almost three-quarters reported recommending CAM therapies in the pre-dissemination survey. Our traditional dissemination methods (print, email, and Internet) were successful in increasing awareness of the CPG in physicians and patients. Overall awareness of the CPG was low in patients despite a statistically significant increase between pre-dissemination and post-traditional dissemination. Traditional dissemination methods were also effective in increasing intent to discuss CAM in both physicians and patients. Despite increased awareness of the CPG, knowledge did not change in physicians across either dissemination method, although an increase in knowledge was noted in CPG-aware physicians as compared with those unaware of the CPG. Knowledge also did not change in patients with either dissemination method, regardless of whether they were aware of the CPG. A significantly greater proportion of CPG-aware physicians and CPG-aware patients reported discussing CAM use with their patients and physicians, respectively, across all the post-dissemination surveys (behavior change).

The lack of change in knowledge among physicians could be due to a ceiling effect in an already enriched population of physicians with interests in CAM and MS and a high baseline level of knowledge of the topic. For patients, the lack of increase in knowledge may have resulted from the overall low awareness. However, knowledge was no different in patients who were aware of the CPG and those who were unaware. It is difficult to draw any conclusions given the small number of aware patients. It is intriguing that intent changed in the absence of increased knowledge. Perhaps the questions used to capture knowledge of the CPG were not directly relevant to intent, which was broadly defined as intent to discuss CAM or intent to start/stop any CAM.

To our surprise, despite an apparently successful dissemination effort using novel media as measured by the reach of Facebook, Twitter, YouTube, and LinkedIn, there was no additional increase in awareness, intent (our primary outcome), knowledge, or behavior in either physicians or patients after social media dissemination efforts as compared with traditional methods. We used targeted advertising to try to reach audiences interested in MS and CAM, and we cast a fairly wide net. The reasons for apparent lack of effectiveness of social media as compared with traditional dissemination methods merit further study but may be related to several factors. First, the CPG may have been considered "old news" when disseminated through social media because traditional dissemination had already saturated the target audience. However, this does not fully explain the large numbers of "hits" that the CPG received on multiple social media networks. Second, despite the fact that the CPG received wide attention through social media advertising, the audience did not follow through by clicking the links to the CPG, and hence, did not remember the CPG or its contents. Third, the respondents of the surveys may not have been users of social media. Finally, social media may not be a useful tool for disseminating CPG recommendations. In a previous study, a social media marketing campaign on public awareness of hypertension did not change knowledge among participants, and the authors suggested that the target of dissemination efforts should be medical professionals in order to increase patient awareness at the point of care [41].

Limitations

Because of the lack of availability of demographics between patient respondents and nonrespondents, we could not evaluate potential confounders. We chose a "between-participants" design rather than a "within-participants" design to avoid the learning and bias that would be expected in serial surveys of the same respondents. Although a comparison of respondents across surveys did not reveal significant differences (data not shown), a difference cannot be excluded with certainty, with resultant effects on our outcome measures.

Finally, contamination between traditional and novel methods cannot be excluded, as we did not have any control over how the CPG, when published, would be shared by the target audience. It is possible that contamination may explain some of the lack of effect of social media. We also recognize that our physician population was an inherently enriched one. An alternative study methodology may be to use a more controlled setting than the real-world setting that we studied. This could potentially be done with a pre- identified study population randomized into two cohorts of traditional dissemination and social media dissemination, with instructions to access the information only through the dissemination method to which they were randomized. However, this would also have limitations, including those of an artifactual setting that may not reflect real-world results. As part of traditional dissemination, a news report of the CPG was published in *Neurology Now*. Because the patient respondents were chosen from the subscriber database of this publication, it is possible that we may have enriched our patient population as well. However, we saw only slight increases in awareness after traditional dissemination, suggesting that patient participants

may not have been enriched. A problem that the AAN anticipated was the ability to control the conversation and the perception of the news media interpretation of the recommendation for use of cannabis extracts. AAN's communications team works with trusted media reporters to promote accurate messaging of the CPG recommendations, to minimize misrepresentation of the evidence, but recognizes that this is not foolproof.

Conclusions

Our results are important in planning future dissemination efforts. Although we did not detect a difference in the effectiveness of the social media–based interventions as compared with the traditional methods, social media were useful in reaching large numbers of the public. We are planning subgroup analyses that may inform future targeted dissemination efforts to those populations that use social media most effectively. The diffusion of innovations model by Rogers is often utilized in disseminations research [42], and it has been suggested that social media platforms such as YouTube, integrated into these models, may be useful [43]. The term "Web 2.0" has been applied to the interactive Internet experience of today [44]. Access, relevance, and credibility have been described as the three critical criteria in using Web 2.0 technologies for dissemination [45]. Partnerships with commercial technology companies, utilization of rapid and adaptive designs to identify successful strategies for user engagement, and iterative evaluation of their efficacy have been recommended to effectively harness Web 2.0 for dissemination [9]. Further research is needed on methods to effectively harness social media platforms that have the potential to easily and inexpensively reach large audiences. This includes audience selection, message formatting and delivery, and other messaging characteristics. Focus group discussions and surveys/interviews of the target audiences (physicians and patients) may provide valuable input to refine social media use in dissemination of CPGs. Research is also needed on milestones and metrics to measure implementation of CPG recommendations.

Conflicts of Interest

SDG has received travel support from the University of Louisville and the New York Academy of Medicine for guideline dissemination talks.

Multimedia Appendix 1

Audience selection for social media dissemination.

[PDF File (Adobe PDF File), 208KB - jmir_v17i8e193_app1.pdf]

Multimedia Appendix 2

Survey development methods and survey data collection procedure.

[PDF File (Adobe PDF File), 70KB - jmir_v17i8e193_app2.pdf]

Multimedia Appendix 3

Surveys used in the study.

[PDF File (Adobe PDF File), 579KB - jmir_v17i8e193_app3.pdf]

Multimedia Appendix 4

Results of dissemination efforts.

[PDF File (Adobe PDF File), 78KB - jmir_v17i8e193_app4.pdf]

References

XSL•FO

RenderX

1. American Academy of Neurology. Clinical Practice Guideline Process Manual. St Paul, MN: The American Academy of Neurology; 2011. URL: https://www.aan.com/Guidelines/Home/Development [accessed 2015-03-13] [WebCite Cache ID 6X0aat9ev]

2. Institute of Medicine of the National Academies. Clinical Practice Guidelines We Can Trust: Standards for Developing Trustworthy Clinical Practice Guidelines. 2011. URL: http://iom.nationalacademies.org/Reports/2011/Clinical-Practice-Guidelines-We-Can-Trust.aspx [accessed 2015-08-07] [WebCite Cache ID 6abwjdhHw]

3. Kryworuchko J, Stacey D, Bai N, Graham ID. Twelve years of clinical practice guideline development, dissemination and evaluation in Canada (1994 to 2005). Implement Sci 2009;4:49 [FREE Full text] [doi: 10.1186/1748-5908-4-49] [Medline: 19656384]

4. Pew Research Center Internet Project. More information yields more learning and sharing. 2014. URL: http://www.pewinternet.org/2014/12/08/more-information-yields-more-learning-and-sharing/ [WebCite Cache ID 6WyaLbaag]

5. Nagendran M, Dimick JB. Disseminating research findings: preparing for Generation Y. JAMA Surg 2014 Jul;149(7):629-630. [doi: 10.1001/jamasurg.2013.5019] [Medline: 24849294]

6. Alas A, Sajadi KP, Goldman HB, Anger JT. The rapidly increasing usefulness of social media in urogynecology. Female Pelvic Med Reconstr Surg 2013;19(4):210-213. [doi: 10.1097/SPV.0b013e3182909872] [Medline: 23797519]

7. Bosley JC, Zhao NW, Hill S, Shofer FS, Asch DA, Becker LB, et al. Decoding twitter: Surveillance and trends for cardiac arrest and resuscitation communication. Resuscitation 2013 Feb;84(2):206-212 [FREE Full text] [doi: 10.1016/j.resuscitation.2012.10.017] [Medline: 23108239]

8. Camm CF, Sunderland N, Camm AJ. A quality assessment of cardiac auscultation material on YouTube. Clin Cardiol 2013 Feb;36(2):77-81. [doi: 10.1002/clc.22080] [Medline: 23172251]

9. Cavallo DN, Chou WS, McQueen A, Ramirez A, Riley WT. Cancer prevention and control interventions using social media: user-generated approaches. Cancer Epidemiol Biomarkers Prev 2014 Sep;23(9):1953-1956. [doi: 10.1158/1055-9965.EPI-14-0593] [Medline: 25103820]

10. Clerici CA, Veneroni L, Bisogno G, Trapuzzano A, Ferrari A. Videos on rhabdomyosarcoma on YouTube: an example of the availability of information on pediatric tumors on the web. J Pediatr Hematol Oncol 2012 Nov;34(8):e329-e331. [doi: 10.1097/MPH.0b013e31825886f8] [Medline: 22858570]

11. Ferguson C, Inglis SC, Newton PJ, Cripps PJS, MacDonald PS, Davidson PM. Social media: a tool to spread information: a case study analysis of twitter conversation at the Cardiac Society of Australia & New Zealand 61st annual scientific meeting 2013. Collegian 2014;21(2):89-93. [Medline: 25109206]

12. Finch CF, McCrory P, Ewing MT, Sullivan SJ. Concussion guidelines need to move from only expert content to also include implementation and dissemination strategies. Br J Sports Med 2013 Jan;47(1):12-14. [doi: 10.1136/bjsports-2012-091796] [Medline: 23060652]

13. George DD, Wainwright BD. Dermatology resources on the internet. Semin Cutan Med Surg 2012 Sep;31(3):183-190. [doi: 10.1016/j.sder.2012.06.002] [Medline: 22929356]

14. Harris JK, Moreland-Russell S, Tabak RG, Ruhr LR, Maier RC. Communication about childhood obesity on Twitter. Am J Public Health 2014 Jul;104(7):e62-e69 [FREE Full text] [doi: 10.2105/AJPH.2013.301860] [Medline: 24832138]

15. Househ M. The use of social media in healthcare: organizational, clinical, and patient perspectives. Stud Health Technol Inform 2013;183:244-248. [Medline: 23388291]

16. Jenkins WD, Wold B. Use of the Internet for the surveillance and prevention of sexually transmitted diseases. Microbes Infect 2012 May;14(5):427-437. [doi: 10.1016/j.micinf.2011.12.006] [Medline: 22207184]

17. Jones K, Baldwin KA, Lewis PR. The potential influence of a social media intervention on risky sexual behavior and Chlamydia incidence. J Community Health Nurs 2012;29(2):106-120. [doi: 10.1080/07370016.2012.670579] [Medline: 22536914]

18. Leveridge MJ. The emerging role of social media in urology. Rev Urol 2014;16(3):110-117. [Medline: 25337040]

19. Mukewar S, Mani P, Wu X, Lopez R, Shen B. YouTube and inflammatory bowel disease. J Crohns Colitis 2013 Jun;7(5):392-402. [doi: 10.1016/j.crohns.2012.07.011] [Medline: 22906403]

20. Savas JA, Huang KE, Tuchayi SM, Feldman SR. Understanding the influence of social media in medicine: lesson learned from Facebook. Dermatol Online J 2014 Sep;20(9):-. [Medline: 25244178]

21. Singh AG, Singh S, Singh PP. YouTube for information on rheumatoid arthritis--a wakeup call? J Rheumatol 2012 May;39(5):899-903. [doi: 10.3899/jrheum.111114] [Medline: 22467934]

22. Syed-Abdul S, Fernandez-Luque L, Jian W, Li Y, Crain S, Hsu M, et al. Misleading health-related information promoted through video-based social media: anorexia on YouTube. J Med Internet Res 2013;15(2):e30 [FREE Full text] [doi: 10.2196/jmir.2237] [Medline: 23406655]

23. Kerber KA, Burke JF, Skolarus LE, Callaghan BC, Fife TD, Baloh RW, et al. A prescription for the Epley maneuver: www.youtube.com? Neurology 2012 Jul 24;79(4):376-380 [FREE Full text] [doi: 10.1212/WNL.0b013e3182604533] [Medline: 22826542]

24. Grande D, Gollust SE, Pany M, Seymour J, Goss A, Kilaru A, et al. Translating research for health policy: researchers' perceptions and use of social media. Health Aff (Millwood) 2014 Jul;33(7):1278-1285. [doi: 10.1377/hlthaff.2014.0300] [Medline: 24907363]

XSL·FO

RenderX

25. Yadav V, Bever C, Bowen J, Bowling A, Weinstock-Guttman B, Cameron M, et al. Summary of evidence-based guideline: complementary and alternative medicine in multiple sclerosis: report of the guideline development subcommittee of the American Academy of Neurology. Neurology 2014 Mar 25;82(12):1083-1092 [FREE Full text] [doi: 10.1212/WNL.0000000000000250] [Medline: 24663230]

26. Schwartz CE, Laitin E, Brotman S, LaRocca N. Utilization of unconventional treatments by persons with MS: is it alternative or complementary? Neurology 1999 Feb;52(3):626-629. [Medline: 10025800]

27. Marrie RA, Hadjimichael O, Vollmer T. Predictors of alternative medicine use by multiple sclerosis patients. Mult Scler 2003 Oct;9(5):461-466. [Medline: 14582770]

28. Nayak S, Matheis RJ, Schoenberger NE, Shiflett SC. Use of unconventional therapies by individuals with multiple sclerosis. Clin Rehabil 2003 Mar;17(2):181-191. [Medline: 12625659]

29. Stuifbergen AK, Harrison TC. Complementary and alternative therapy use in persons with multiple sclerosis. Rehabil Nurs 2003;28(5):141-7, 158. [Medline: 14521002]

30. Apel A, Greim B, Zettl UK. How frequently do patients with multiple sclerosis use complementary and alternative medicine? Complement Ther Med 2005 Dec;13(4):258-263. [doi: 10.1016/j.ctim.2005.05.006] [Medline: 16338196]

31. Apel A, Greim B, König N, Zettl UK. Frequency of current utilisation of complementary and alternative medicine by patients with multiple sclerosis. J Neurol 2006 Oct;253(10):1331-1336. [doi: 10.1007/s00415-006-0217-9] [Medline: 16786211]

32. Schwarz S, Knorr C, Geiger H, Flachenecker P. Complementary and alternative medicine for multiple sclerosis. Mult Scler 2008 Sep;14(8):1113-1119. [doi: 10.1177/1352458508092808] [Medline: 18632773]

33. Pucci E, Cartechini E, Taus C, Giuliani G. Why physicians need to look more closely at the use of complementary and alternative medicine by multiple sclerosis patients. Eur J Neurol 2004 Apr;11(4):263-267. [doi: 10.1046/j.1468-1331.2003.00758.x] [Medline: 15061828]

34. Marrie RA, Salter AR, Tyry T, Fox RJ, Cutter GR. Preferred sources of health information in persons with multiple sclerosis: degree of trust and information sought. J Med Internet Res 2013;15(4):e67 [FREE Full text] [doi: 10.2196/jmir.2466] [Medline: 23635393]

35. Schillinger D. An introduction to effectiveness, dissemination and implementation research. University of California: From the series: UCSF clinical and translational science institute (CTSI) resource manuals and guides to community- engaged research. Fleischer P (ed) Published by Clinical Translational Science Institute Community Engagement Program, URL: https://accelerate.ucsf.edu/files/CE/edi_introguide.pdf [accessed 2015-07-12] [WebCite Cache ID 6ZyLeLjRn]

36. Albrecht L, Archibald M, Arseneau D, Scott SD. Development of a checklist to assess the quality of reporting of knowledge translation interventions using the Workgroup for Intervention Development and Evaluation Research (WIDER) recommendations. Implement Sci 2013;8:52 [FREE Full text] [doi: 10.1186/1748-5908-8-52] [Medline: 23680355]

37. American Association of Public Opinion Research. Best practices for research URL: http://www.aapor.org/AAPORKentico/Standards-Ethics/Best-Practices.aspx [accessed 2015-03-12] [WebCite Cache ID 6WyaZZ8IP]

38. Kuhl J, Beckmann J. A theory of planned behavior. In: Action control, from cognition to behavior. Berlin: Springer-Verlag; 1985:11-39.

39. Cabana MD, Rand CS, Powe NR, Wu AW, Wilson MH, Abboud PA, et al. Why don't physicians follow clinical practice guidelines? A framework for improvement. JAMA 1999 Oct 20;282(15):1458-1465. [Medline: 10535437]

40. National Information Center on Heath Services Research and Health Care Technology. URL: http://www.nlm.nih.gov/hsrinfo/implementation_science.html [accessed 2015-03-12] [WebCite Cache ID 6WyadWzYV]

41. Petrella RJ, Speechley M, Kleinstiver PW, Ruddy T. Impact of a social marketing media campaign on public awareness of hypertension. Am J Hypertens 2005 Feb;18(2 Pt 1):270-275. [doi: 10.1016/j.amjhyper.2004.09.012] [Medline: 15752956]

42. Rogers E. Diffusion of innovations. New York: Free Press; 2003.

43. Lillie SE. Diffusion of innovation in the age of YouTube. Am J Prev Med 2008 Mar;34(3):267. [doi: 10.1016/j.amepre.2007.11.009] [Medline: 18312819]

44. O'Reilly T. What is Web 2.0: design patterns and business models for the next generation of software. Sebastopol, CA: O'Reilly Media, Inc; 2005. URL: http://www.oreilly.com/pub/a/web2/archive/what-is-web-20.html [accessed 2015-03-12] [WebCite Cache ID 6WyahmtCL]

45. Metzger MJ, Flanagin AJ. Using Web 2.0 technologies to enhance evidence-based medical information. J Health Commun 2011;16 Suppl 1:45-58. [doi: 10.1080/10810730.2011.589881] [Medline: 21843095]

Abbreviations

AAN: American Academy of Neurology
AD: absolute difference
CAM: complementary and alternative medicine
"CAM in MS": "Complementary and alternative medicine in multiple sclerosis"
CPG: clinical practice guideline
HMO: health maintenance organization

MS: multiple sclerosis
NMSS: National Multiple Sclerosis Society

XSL•FO
RenderX

Original Paper

Predictors of "Liking" Three Types of Health and Fitness-Related Content on Social Media: A Cross-Sectional Study

Elise R Carrotte[1], B Psych (Hons); Alyce M Vella[1], BSc (Hons); Megan SC Lim[1,2], PhD

[1]Centre for Population Health, Burnet Institute, Melbourne, Australia

[2]School of Public Health and Preventive Medicine, Monash University, Melbourne, Australia

Corresponding Author:
Elise R Carrotte, B Psych (Hons)
Centre for Population Health
Burnet Institute
85 Commercial Road
Melbourne, 3004
Australia
Phone: 61 385062365
Fax: 61 392822100
Email: elise.carrotte@burnet.edu.au

Abstract

Background: Adolescence and young adulthood are key periods for developing norms related to health behaviors and body image, and social media can influence these norms. Social media is saturated with content related to dieting, fitness, and health. Health and fitness–related social media content has received significant media attention for often containing objectifying and inaccurate health messages. Limited research has identified problematic features of such content, including stigmatizing language around weight, portraying guilt-related messages regarding food, and praising thinness. However, no research has identified who is "liking" or "following" (ie, consuming) such content.

Objective: This exploratory study aimed to identify demographics, mental health, and substance use–related behaviors that predicted consuming 3 types of health and fitness–related social media content—weight loss/fitness motivation pages (ie, "fitspiration"), detox/cleanse pages, and diet/fitness plan pages—among young social media users.

Methods: Participants (N=1001; age: median 21.06, IQR 17.64-24.64; female: 723/1001, 72.23%) completed a cross-sectional 112-question online survey aimed at social media users aged between 15-29 years residing in Victoria, Australia. Logistic regression was used to determine which characteristics predicted consuming the 3 types of health and fitness–related social media content.

Results: A total of 378 (37.76%) participants reported consuming at least 1 of the 3 types of health and fitness–related social media content: 308 (30.77%) fitspiration pages, 145 (14.49%) detox pages, and 235 (23.48%) diet/fitness plan pages. Of the health and fitness–related social media content consumers, 85.7% (324/378) identified as female and 44.8% (324/723) of all female participants consumed at least 1 type of health and fitness–related social media content. Predictors of consuming at least one type of health and fitness–related social media content in univariable analysis included female gender (OR 3.5, 95% CI 2.5-4.9, *P*<.001), being aged 15-17 years (OR 3.0, 95% CI 2.2-4.0, *P*<.001), residing outside a major city (OR 2.0, 95% CI 1.4-2.9, *P*<.001), having no post–high school education (OR 2.2, 95% CI 1.7-2.9, *P*<.001), being born in Australia (OR 2.0, 95% CI 1.2-3.2, *P*=.006), having a self-reported eating disorder (OR 2.4, 95% CI 1.5-3.9, *P*<.001), being a victim of bullying (OR 1.7, CI 1.3-2.3, *P*<.001), misusing detox/laxative teas or diet pills (OR 4.6, 95% CI 2.8-7.6, *P*<.001), never using illegal drugs (OR 1.6, 95% CI 1.2-2.0, *P*=.001), and not engaging in risky single occasion drinking on a weekly basis (OR 2.0, 95% CI 1.3-3.0, *P*=.003).

Conclusions: Consumers of health and fitness–related social media content were predominantly teenaged girls. There is a need to ensure that this social media content portrays responsible health messages and to research further the role of fitspiration pages, detox pages, and diet/fitness plan pages in influencing body image and health behaviors.

Introduction

Background

Social media is widely used and accepted among young people. In the United States, up to 90% of teenagers and young adults report using Facebook, whereas more than half use Instagram and one-third use Twitter [1,2]. Young people are increasingly turning to social media as a source of health-related information [3]. A plethora of health and fitness-related social media content is available to young people and is popular, diverse, and interactive; when social media users "like" or "follow" health and fitness-related social media content pages, content appears in their newsfeeds where the user can view and engage with the content by commenting on photos or sharing with friends (through "tagging" or reposting content). One type of health and fitness-related social media content, "fitspiration," refers to messages designed to inspire individuals to achieve a health or fitness goal, usually through exercise and dieting [4]. Common forms of fitspiration include images of toned bodies overlaid with quotes designed to motivate viewers (Figure 1), blog entries, and personal stories (eg, "before-and-after" weight loss pictures), and personal profiles of fitness trainers and fitness models. Other forms of health and fitness-related social media content include strict diet/exercise plans and "cleanses" or "detoxes" that claim to have health and weight loss benefits.

Health and fitness-related social media content is commonly posted by companies to sell a service or product (eg, personal trainers, gyms, or brands of juice detoxes). Health and fitness-related social media content can also be user-generated and maintained; for example, some social media users commonly post exercise "selfies" (self-portrait photographs), statuses about fitness routines, and images of healthy food desired or prepared by the user [3,5].

Health and fitness-related social media content appears to be a double-edged sword. Social media can play a role in shaping body image through social comparison with others and the maintenance of weight- and appearance-related concerns [6-8]. For example, in an exploratory qualitative study into social media's influence on health behaviors [3], young American adults (mean age 20.4 years) agreed that seeing exercise tips and instructions, using exercise tracking apps, and viewing weight loss before-and-after pictures and fitness-related quotes can be motivational for improving health behaviors. However, some content, such as friends posting fitness-related selfies with negative captions about their physical appearance (eg, "[I'm] still really fat") can induce negative feelings of body-related shame in the viewer. Content may be misleading, such as advertisements or fitness programs/products conveying unrealistic goals, and users of social media often wish to look their best for their social network and are selective about the content they post [3].

Figure 1. Examples of fitspiration-style images. Photo credit: Shutterstock.

Criticisms of Health and Fitness–Related Social Media Content

Although many social media-based health and fitness initiatives and interventions are based on scientific research and are run by qualified teams of health experts [9], some health and fitness-related social media content has been criticized for sending inaccurate or irresponsible health messages, a topic of recent media debate. Common criticisms of fitspiration include the prominence of fitspiration images that champion pushing oneself too far during exercise, focus on appearance rather than fitness, and praise the athletic body type (the "athletic ideal") [10] (Figure 1). Internalization of the athletic ideal has been associated with increased compulsive exercising and negative mood associated with missing an exercise session [11]. Despite a general focus on the athletic ideal body, fitspiration aimed at women often relies on images of slim or thin female bodies to promote an image of what it means to be healthy, fit, and strong

[12]. As such, fitspiration has been compared to "thinspiration" and "pro-ana" (pro-anorexia/eating disorder) content, which idealizes thin bodies (the "thin ideal" [13]) and is designed to motivate viewers to lose weight. Exposure to these websites has been associated with adverse effects, such as negative mood and lowered self-esteem, decreased perceived attractiveness, and increased dieting in experimental studies [14,15]. Meanwhile, more than one-third of young people with eating disorders have reported visiting these sites and learning new weight loss and purging techniques [16].

Forms of health and fitness-related social media content that focus on diet, health, and well-being have also been criticized. For example, 2 popular diet programs—the Paleo Diet and the Sugar Free Diet—have been listed by the British Dietetic Association as two of the "Worst Celebrity Diets" and criticized for being unbalanced and unnecessarily restrictive of food groups [17]. Similarly, liquid-based detox diets that claim to rid the body of toxins (despite no medical evidence indicating

this is necessary) have been criticized by the Dietitians Association of Australia and can result in the loss of healthy gut bacteria and electrolytes [18]. Despite these concerns, 42% of American adult social media users have reported that information found via social media would affect health decisions related to diet, exercise, or stress management, and nearly 90% of people aged 18 to 24 years have indicated they would trust medical information found on social media [19].

Past Research

Two recent analyses have indicated that problematic content is posted regularly online under the guise of health. In the first study, which analyzed 21 "healthy living" blogs (which aim to offer advice and personal experiences regarding health), it was found that approximately half contained content with negative or guilt-inducing messages about food and/or content with stigmatizing language relating to weight [20]. In the second study, the authors compared 50 websites dedicated to fitspiration with 50 thinspiration websites. The authors found that although thinspiration websites were more likely than fitspiration websites to contain content praising thinness (34% vs 10%) and championing weight loss (68% vs 42%), both types of websites were equally as likely to contain objectifying content (32% vs 36%), guilt-inducing messages about weight or the body (both 36%), and stigmatizing messages around fat and weight (both 20%) [4]. The authors suggested that although thinspiration appears more obviously detrimental to viewers' health and body image, it is within reason to assume that viewing both types of websites may negatively impact viewers [4]. Therefore, fitspiration may also attract people with preexisting eating disorder symptomology or vulnerability, or influence emerging psychological concerns such as orthorexia nervosa, an obsession with healthy eating, food quality, and food "purity" with links to obsessive-compulsive disorder and anorexia [21].

This Study

Despite significant debate in the media about the potential harms of health and fitness-related social media content [12,17,18], little research has examined this content. Previous research has focused on the content of self-labeled fitspiration rather than expanding the scope of this field of research to observe other types of health and fitness-related social media content. Specifically, it is unclear who is liking and following ("consuming") health and fitness-related social media content via social media. The overarching aim of this exploratory study was to identify the characteristics of young people who consume 3 types of health and fitness-related social media content: fitspiration pages, detox pages, and diet/fitness plan pages. The secondary aims of this study were to determine (1) which demographics predict consuming health and fitness-related social media content, (2) whether young people who consume health and fitness-related social media content have poorer self-rated mental health than those who do not consume this content, and (3) whether young people who consume health and fitness-related social media content use various legal and illegal substances at different rates compared to those who do not consume this content. It was hypothesized that health and fitness-related social media content would be more commonly consumed by young women versus young men and that

self-reported mental health problems and misuse of detox teas/laxatives and diet pills would be associated with consuming health and fitness-related social media content.

Methods

Data

Participants were recruited via the 2015 Sex, Drugs and Rock'n'Roll study, developed by Burnet Institute: a cross-sectional convenience sample of people aged 15 to 29 years living in Victoria, Australia. The study consisted of a 112-question online survey, which covered demographics, social media use, and general mental and sexual health. The survey was available on Burnet Institute's website for 6 weeks between February and March 2015. Participants were recruited via social media, advertisements on Facebook targeted to young people, and word of mouth. Only complete responses were analyzed. Participants had the opportunity to win a gift voucher for participating. Informed consent was obtained from each participant. Approval for this study was granted by the Alfred Hospital Human Research Ethics Committee, Melbourne, Victoria. No specific funding was received for this study.

Measures

Demographics

Demographic details included gender (male, female, transgender, or other with option to specify) and age, which was calculated from month and year of birth. A binary variable was created for gender (male/female; due to sample size, only participants identifying as male and female were included in gender-related data analyses) and age was recoded into 3 categories (15-17, 18-19, and 20-29 years). Country of birth was dichotomized as Australian-born or born outside Australia. Participants specified their highest level of completed education, which informed a binary variable which distinguished between participants who were currently completing or had completed any post–high school education and those with high school education or lower (including those still at high school). Participants indicated their sexual identity; a binary variable distinguished between participants identifying as heterosexual and participants identifying as gay, homosexual, lesbian, bisexual, queer, questioning, or other (GLBQQ+). Recreational spending was analyzed to assess socioeconomic status and dichotomized as less than AUD $120 to spend on oneself per week or AUD $120 or more. Participants' postcodes informed a binary variable to indicate their area of residence, which was major city or nonmajor city.

Health and Fitness–Related Social Media Content

Participants were asked, "Do you like/follow any of the following types of pages on Facebook, Instagram, or Twitter?" with the option to choose all that applied. Four of the options were as follows, based on researcher observations of 3 common types of health and fitness–related social media content and a fourth option that was used for the purpose of comparison:

1. Weight loss/fitness motivation profiles (eg, personal trainers, athletes, fitness models)
2. Cleanses or detoxes (eg, , juice detox)

3. Diet plans or weight loss/fitness challenges (eg, I Quit Sugar, Michelle Bridges 12WBT, Kayla Itsines Bikini Body Challenge)
4. Other health-related pages (eg, Cancer Council)

Binary variables were created to identify whether participants liked or followed (ie, consumed) the 3 types of health and fitness–related social media content of interest (hereafter referred to as fitspiration pages, detox pages, and diet/fitness plan pages) and other health pages, respectively (see Multimedia Appendices 1-3 for screenshot examples). A binary yes/no variable was created to identify whether participants consumed at least 1 of the 3 types of health and fitness–related social media content of interest.

Mental Health

Participants were asked, "In the last 6 months have you had any mental health problems? This includes any issues that you haven't spoken to a health professional about." Options were "yes," "no," or "I don't wish to say." If participants answered "yes" to having mental health problems in the last 6 months, they were asked, "Could you please specify what this mental health problem/s was?" with an option to choose all that applied. Options were anxiety disorder (eg, generalized anxiety disorder, obsessive-compulsive disorder), mood disorder (eg, depression, bipolar disorder), eating disorder (eg, anorexia nervosa, bulimia), "I don't wish to say," and "other" with the option to specify. Binary yes/no variables were created to identify participants experiencing anxiety disorders, mood disorders, and/or eating disorders. Participants were asked if they had been the victim of bullying in the last 6 months, which informed a binary yes/no variable.

Substance Use

Participants were asked to report illegal drug use both in their lifetime and in the last month, if they currently smoked cigarettes, and how often they consumed alcohol. Binary yes/no variables were created for these behaviors; weekly "risky single occasion drinking" was defined as consuming 6 or more standard drinks on a weekly basis based on the Alcohol Use Disorders Identification Test [22]. Participants were also asked, "'In the last 12 months, have you used any of the following drugs/substances illicitly, not as directed or prescribed to someone else? (Tick all that apply)" with options including diet pills and detox/laxative teas. A binary yes/no variable was created to identify participants who had ever misused either diet pills or detox/laxative teas.

Analysis

All statistical analyses were performed using Stata version 13 (StataCorp LP, College Station, TX, USA). Cross-tabulations and univariable logistic regression were used to compare differences in demographics, mental health, and substance use between those who consumed health and fitness–related social media content ("consumers") and those who did not. Multivariable logistic regression was performed using variables significant at $P<.05$ at the univariable level to identify independent predictors of consuming health and fitness–related social media content. In the multivariable model, variables significant with the Bonferroni-adjusted P value of .0125 (.05/4 tests) were deemed to be significant independent predictors of consuming health and fitness–related social media content.

Results

The survey was completed by 1001 participants. The mean age was 21.40 years (SD 4.12) and the median age was 21.06 years (IQR 17.64-24.74); 269 (26.87%) identified as male, 723 (72.23%) identified as female, 4 (0.40%) identified as transgender, 3 (0.20%) reported their gender as "other," and 2 (0.20%) did not specify their gender. A total of 308 participants (30.77%) consumed fitspiration pages, 145 (14.49%) consumed detox pages, and 235 (23.48%) consumed diet/fitness plan pages. In all, 378 (37.76%) participants consumed at least 1 of the 3 types of health and fitness–related social media content, 212 (21.17%) consumed at least 2 types, and 96 (9.59%) participants consumed all 3 types. Of the health and fitness–related social media content consumers, 85.7% (324/378) identified as female and 44.8% (324/723) of all female participants consumed at least one type of health and fitness–related social media content. Further, 57.1% (184/322) of teenaged girls consumed at least one type of health and fitness-related social media content; 48.7% (184/378) of all consumers were teenaged girls.

Univariable logistic regression compared health and fitness–related social media content consumers (378/1001, 37.76%) and consumers of other health pages (358/1001, 35.76%). Consuming other health pages predicted consuming health and fitness–related social media content (OR 2.6, 95% CI 2.0-3.4, $P<.001$). Other health pages were significantly more likely to be consumed by female participants than male participants (OR 1.6, 95% CI 1.2-2.1, $P=.003$) and GLBQQ+ participants than heterosexual participants (OR 1.6, 95% CI 1.2-2.1, $P=.003$). No significant results were observed for other demographics regarding other health pages.

Logistic regression was used to examine correlates of consuming at least 1 of the 3 types of health and fitness–related social media content (Table 1). In univariable analysis, significant differences ($P<.05$) were found; consumers of any health and fitness–related social media content were more likely to report female gender, younger age, location in a nonmajor city, no post–high school education, being born in Australia, experiencing eating disorders, being a victim of bullying, misusing detox/laxative teas or diet pills, never using illegal drugs, and not engaging in weekly risky single occasion drinking compared to those who did not consume any health and fitness–related social media content. In multivariable analysis (pseudo R^2=.11), significant independent predictors of consuming any health and fitness–related social media content at Bonferroni-adjusted $P<.0125$ were female gender (OR 2.6, 95% CI 1.8-3.7, $P<.001$), being aged 15-17 years (OR 2.5, 95% CI 1.4-4.4, $P=.002$), and misusing diet pills or detox teas (OR 3.5, 95% CI 2.0-5.9, $P<.001$).

Table 1. Descriptive statistics and univariable logistic regression comparing consumers of at least one type of health and fitness–related social media content (consumers) and participants who did not consume any health and fitness–related social media content (nonconsumers).

Variable	Total, n (%) N=1001	Consumers, n (%) n=378	Nonconsumers, n (%) n=623	OR (95% CI)	P
Gender					
Male	269 (26.87)	51 (19.0)	218 (81.0)	1.0	
Female	723 (72.23)	324 (44.8)	399 (55.2)	3.5 (2.5-4.9)	<.001
Age (years)					
15-17	279 (27.87)	154 (55.2)	125 (44.8)	3.0 (2.2-4.0)	<.001
18-19	128 (12.79)	51 (39.8)	77 (60.2)	1.6 (1.1-2.4)	.02
20-29	594 (59.34)	173 (29.1)	421 (70.9)	1.0	
Place of residence					
Nonmajor city	137 (13.69)	72 (52.6)	65 (47.5)	2.0 (1.4-2.9)	<.001
Major city	846 (84.52)	300 (35.6)	246 (64.5)	1.0	
Education					
No post–high school	336 (33.57)	169 (50.3)	167 (49.7)	2.2 (1.7-2.9)	<.001
Post–high school	664 (66.33)	208 (31.3)	456 (68.7)	1.0	
Country of birth					
Australia	893 (89.21)	350 (39.2)	543 (60.8)	2.0 (1.2-3.2)	.006
Outside Australia	97 (9.69)	24 (24.7)	73 (75.3)	1.0	
Sexual identity					
Heterosexual	768 (76.72)	301 (39.2)	467 (60.8)	1.3 (1.0-1.8)	.10
GLBQQ+[a]	229 (22.88)	76 (33.2)	153 (66.8)	1.0	
Recreational spending per week (AUD $)					
<$120	770 (76.92)	299 (38.8)	471 (61.2)	1.2 (0.9-1.7)	.20
≥$120	226 (22.58)	77 (34.1)	149 (65.9)	1.0	
Anxiety [b]					
Yes	404 (40.36)	156 (38.6)	248 (61.4)	1.1 (0.8-1.4)	.640
No	597 (59.64)	222 (37.2)	375 (62.8)	1.0	
Eating disorder [b]					
Yes	76 (7.59)	44 (57.9)	32 (42.1)	2.4 (1.5-3.9)	<.001
No	925 (92.41)	334 (36.1)	591 (63.9)	1.0	
Mood disorder [b]					
Yes	365 (36.46)	151 (41.4)	214 (58.6)	1.3 (1.0-1.7)	.08
No	636 (63.54)	227 (35.7)	709 (64.3)	1.0	
Bullied (last 6 months)					
Yes	265 (26.47)	126 (47.6)	139 (52.5)	1.7 (1.3-2.3)	<.001
No	736 (73.53)	252 (34.2)	484 (65.8)	1.0	
Misused detox/laxative teas or diet pills (last 12 months)					
Yes	83 (8.29)	59 (71.1)	24 (28.9)	4.6 (2.8-7.6)	<.001
No	918 (91.71)	319 (34.8)	599 (65.3)	1.0	
Ever used illegal drugs					
Yes	545 (54.45)	180 (33.0)	365 (67.0)	1.0	
No	446 (44.56)	195 (43.7)	257 (56.3)	1.6 (1.2-2.0)	.001

Variable	Total, n (%) N=1001	Consumers, n (%) n=378	Nonconsumers, n (%) n=623	OR (95% CI)	P
Last month illegal drug use if ever used drugs					
Yes	330 (33.00)	111 (33.6)	219 (66.4)	1.1 (0.7-1.5)	.71
No	215 (21.48)	69 (32.1)	146 (67.9)	1.0	
Risky single occasion drinking weekly or more often					
Yes	122 (12.19)	30 (24.6)	92 (75.4)	1.0	
No	758 (75.72)	295 (38.9)	61.1 (463)	2.0 (1.3-3.0)	.003
Current smoker					
Yes	206 (20.68)	71 (34.6)	135 (65.6)	1.0	
No	790 (79.32)	304 (38.5)	486 (61.5)	1.2 (0.9-1.6)	.290

[a] Gay, lesbian, bisexual, queer, or questioning.

[b] Based on self-reported diagnosed and undiagnosed conditions in the last 6 months.

Univariable analyses were repeated for each type of health and fitness–related social media content separately (Tables 2-4). In multivariable regression (pseudo R^2=.09), significant independent predictors of consuming fitspiration pages at Bonferroni-adjusted $P<.0125$ were female gender (OR 2.0, 95% CI 1.4-2.8, $P<.001$), being aged 15-17 years (OR 2.7, 95% CI 1.5-4.9, $P=.002$), identifying as heterosexual (OR 1.6, 95% CI 1.1-2.4, $P=.009$), and misusing diet pills or detox teas (OR 2.1, 95% CI 1.3-3.5, $P=.004$). Significant independent predictors of consuming detox pages at Bonferroni-adjusted $P=.01$ in

multivariable regression (pseudo R^2=.24) were female gender (OR 52.1, 95% CI 7.2-377.6, $P<.001$), being aged 15-17 years (OR 3.4, 95% CI 1.5-7.7, $P=.003$), misusing diet pills or detox teas (OR 4.7, 95% CI 2.7-8.0, $P<.001$) and using illegal drugs in the last month (OR 2.5, 95% CI 1.3-4.9, $P=.008$). Significant independent predictors of consuming diet/fitness plan pages at Bonferroni-adjusted $P<.0125$ in multivariable regression (pseudo R^2=.17) were female gender (OR 9.8, 95% CI 4.9-19.6, $P<.001$) and misusing diet pills or detox teas (OR 3.6, 95% CI 2.2-6.0, $P<.001$).

Table 2. Descriptive statistics and univariable logistic comparing consumers of fitspiration pages and nonconsumers.

Variable	Total, n (%) N=1001	Consumers, n (%) n=308	Nonconsumers, n (%) n=693	OR (95% CI)	P
Gender					
Male	269 (26.87)	48 (17.8)	221 (82.2)	1.0	
Female	723 (72.23)	259 (35.8)	464 (64.2)	2.6 (1.8-3.6)	<.001
Age (years)					
15-17	279 (27.87)	128 (45.9)	151 (54.1)	2.8 (2.1-3.8)	<.001
18-19	128 (12.79)	43 (33.6)	85 (66.4)	1.7 (1.1-2.6)	.01
20-29	594 (59.34)	137 (23.1)	457 (76.9)	1.0	
Location					
Nonmajor city	137 (13.69)	60 (43.8)	77 (65.2)	1.9 (1.3-2.8)	<.001
Major city	846 (84.52)	243 (28.7)	603 (71.3)	1.0	
Education					
No post–high school	336 (33.57)	138 (41.1)	198 (58.9)	2.0 (1.5-2.7)	<.001
Post–high-school	664 (66.33)	169 (25.5)	495 (74.6)	1.0	
Country of birth					
Australia	893 (89.21)	286 (32.0)	607 (68.0)	1.8 (1.1-3.0)	.02
Outside Australia	97 (9.69)	20 (20.6)	77 (79.4)	1.0	
Sexual identity					
Heterosexual	768 (76.72)	251 (32.7)	517 (67.3)	1.5 (1.1-2.1)	.02
GLBQQ+[a]	229 (22.88)	56 (24.5)	173 (75.6)	1.0	
Recreational spending per week (AUD $)					
<$120	770 (76.92)	248 (32.3)	522 (67.8)	1.3 (1.0-1.9)	
≥$120	226 (22.58)	59 (26.1)	167 (73.9)	1.0	.08
Anxiety [b]					
Yes	404 (40.36)	125 (30.9)	279 (69.1)	1.0 (0.8-1.3)	.92
No	597 (59.64)	183 (30.7)	414 (69.3)	1.0	
Eating disorder [b]					
Yes	76 (7.59)	38 (50.0)	38 (50.0)	2.4 (1.5-3.9)	<.001
No	925 (92.41)	270 (29.2)	655 (70.8)	1.0	
Mood disorder [b]					
Yes	365 (36.46)	124 (34.0)	241 (66.0)	1.3 (1.0-1.7)	.10
No	636 (63.54)	184 (28.9)	452 (71.1)	1.0	
Bullied (last 6 months)					
Yes	265 (26.47)	100 (37.7)	165 (62.3)	1.5 (1.1-2.1)	.004
No	736 (73.53)	208 (28.3)	528 (71.7)	1.0	
Misused detox/laxative teas or diet pills (last 12 months)					
Yes	83 (8.29)	44 (53.0)	39 (47.0)	2.8 (1.8-4.4)	<.001
No	918 (91.71)	264 (28.8)	654 (71.2)	1.0	
Ever used illegal drugs					
Yes	545 (54.45)	140 (25.7)	405 (74.3)	1.0	
No	446 (44.56)	166 (37.2)	280 (62.8)	1.7 (1.3-2.3)	<.001

Variable	Total, n (%) N=1001	Consumers, n (%) n=308	Nonconsumers, n (%) n=693	OR (95% CI)	*P*
Last month illegal drug use if ever used drugs					
Yes	330 (33.00)	84 (25.5)	246 (74.6)	0.9 (0.7-1.4)	.88
No	215 (21.48)	56 (26.0)	159 (74.0)	1.0	
Risky single occasion drinking weekly or more often					
Yes	122 (12.19)	240 (31.7)	518 (68.3)	1.0	
No	758 (75.72)	21 (17.2)	191 (82.8)	2.2 (1.4-3.7)	.001
Current smoker					
Yes	206 (20.68)	58 (28.1)	148 (71.8)	1.0	
No	790 (79.32)	247 (31.3)	543 (68.7)	1.2 (0.8-1.6)	.39

[a] Gay, lesbian, bisexual, queer, or questioning.

[b] Based on self-reported diagnosed and undiagnosed conditions in the last 6 months.

Table 3. Descriptive statistics and univariable logistic comparing consumers of detox pages and nonconsumers.

Variable	Total, n (%) N=1001	Consumers, n (%) n=145	Nonconsumers, n (%) n=856	OR (95% CI)	P
Gender					
Male	269 (26.87)	1 (0.4)	268 (99.6)	1.0	
Female	723 (72.23)	142 (19.6)	581 (80.4)	65.5 (9.1-470.7)	<.001
Age (years)					
15-17	279 (27.87)	81 (29.0)	198 (71.0)	5.2 (3.5-7.9)	<.001
18-19	128 (12.79)	21 (16.4)	107 (83.6)	2.5 (1.4-4.4)	.001
20-29	594 (59.34)	43 (7.2)	551 (92.8)	1.0	
Location					
Nonmajor city	137 (13.69)	36 (26.3)	101 (73.7)	2.4 (1.6-3.7)	<.001
Major city	846 (84.52)	108 (12.8)	738 (87.2)	1.0	
Education					
No post–high school	336 (33.57)	89 (26.5)	247 (73.5)	3.9 (2.7-5.6)	<.001
Post–high school	664 (66.33)	56 (8.4)	608 (91.6)	1.0	
Country of birth					
Australia	893 (89.21)	136 (15.2)	757 (84.8)	2.0 (0.9-4.2)	.07
Outside Australia	97 (9.69)	8 (8.3)	89 (91.8)	1.0	
Sexual identity					
Heterosexual	768 (76.72)	118 (15.4)	650 (84.6)	1.4 (0.9-2.2)	.13
GLBQQ+[a]	229 (22.88)	26 (11.4)	203 (88.7)	1.0	
Recreational spending per week (AUD $)					
<$120	770 (76.92)	122 (15.8)	648 (84.2)	1.7 (1.1-2.8)	.02
≥$120	226 (22.58)	22 (9.7)	204 (90.3)	1.0	
Anxiety [b]					
Yes	404 (40.36)	67 (16.6)	337 (83.4)	1.3 (0.9-1.9)	.12
No	597 (59.64)	78 (13.1)	519 (86.9)	1.0	
Eating disorder [b]					
Yes	76 (7.59)	25 (32.9)	51 (67.1)	3.3 (2.0-5.5)	<.001
No	925 (92.41)	120 (13.0)	805 (87.0)	1.0	
Mood disorder [b]					
Yes	365 (36.46)	63 (17.3)	302 (82.7)	1.4 (1.0-2.0)	.06
No	636 (63.54)	82 (12.9)	553 (87.1)	1.0	
Bullied (last 6 months)					
Yes	265 (26.47)	51 (19.3)	214 (80.8)	1.6 (1.1-2.4)	.01
No	736 (73.53)	94 (12.8)	642 (87.2)	1.0	
Misused detox/laxative teas or diet pills (last 12 months)					
Yes	83 (8.29)	40 (48.2)	43 (51.8)	7.2 (4.5-11.6)	<.001
No	918 (91.71)	105 (11.4)	813 (88.6)	1.0	
Ever used illegal drugs					
Yes	545 (54.45)	72 (13.2)	473 (86.8)	1.0	
No	446 (44.56)	72 (16.1)	374 (83.9)	1.3 (0.9-1.8)	.19

XSL•FO
RenderX

Variable	Total, n (%) N=1001	Consumers, n (%) n=145	Nonconsumers, n (%) n=856	OR (95% CI)	*P*
Last month illegal drug use if ever used drugs					
Yes	330 (33.00)	57 (17.3)	273 (82.7)	2.8 (1.5-5.1)	.001
No	215 (21.48)	15 (7.0)	200 (93.0)	1.0	
Risky single occasion drinking weekly or more often					
Yes	122 (12.19)	12 (9.8)	110 (90.2)	1.0	
No	758 (75.72)	112 (14.8)	646 (85.2)	1.6 (0.9-3.0)	.15
Current smoker					
Yes	206 (20.68)	37 (18.0)	169 (82.0)	1.4 (0.9-2.1)	.10
No	790 (79.32)	106 (13.4)	684 (86.6)	1.0	

[a] Gay, lesbian, bisexual, queer, or questioning.

[b] Based on self-reported diagnosed and undiagnosed conditions in the last 6 months.

Table 4. Descriptive statistics and univariable logistic comparing consumers of diet/fitness plan pages and nonconsumers.

Variable	Total, n (%) N=1001	Consumers, n (%) n=235	Nonconsumers, n (%) n=766	OR (95% CI)	*P*
Gender					
Male	269 (26.87)	9 (3.4)	260 (96.7)	1.0	
Female	723 (72.23)	225 (31.1)	498 (68.9)	13.0 (6.6-25.8)	<.001
Age (years)					
15-17	279 (27.87)	106 (38.0)	173 (62.0)	3.2 (2.3-4.4)	<.001
18-19	128 (12.79)	33 (25.8)	95 (74.2)	1.8 (1.1-2.8)	.01
20-29	594 (59.34)	96 (16.2)	498 (83.8)	1.0	
Location					
Nonmajor city	137 (13.69)	40 (29.2)	97 (70.8)	1.4 (0.9-2.1)	.09
Major city	846 (84.52)	191 (22.6)	655 (77.4)	1.0	
Education					
No post–high school	336 (33.57)	118 (35.1)	218 (64.9)	2.5 (1.9-3.4)	<.001
Post–high school	664 (66.33)	117 (17.6)	547 (82.4)	1.0	
Country of birth					
Australia	893 (89.21)	221 (24.8)	672 (75.3)	2.3 (1.2-4.3)	.008
Outside Australia	97 (9.69)	12 (12.4)	85 (87.6)	1.0	
Sexual identity					
Heterosexual	768 (76.72)	189 (24.6)	579 (75.4)	1.3 (0.9-1.9)	.12
GLBQQ+[a]	229 (22.88)	45 (19.7)	184 (80.4)	1.0	
Recreational spending per week (AUD $)					
<$120	770 (76.92)	190 (24.7)	580 (75.3)	1.4 (1.0-2.0)	.08
≥$120	226 (22.58)	43 (19.0)	183 (81.0)	1.0	
Anxiety [b]					
Yes	404 (40.36)	107 (26.5)	297 (73.5)	1.3 (1.0-1.8)	.07
No	597 (59.64)	128 (21.4)	469 (78.6)	1.0	
Eating disorder [b]					
Yes	76 (7.59)	35 (46.0)	41 (54.0)	3.1 (1.9-5.0)	<.001
No	925 (92.41)	200 (31.6)	725 (78.4)	1.0	
Mood disorder [b]					
Yes	365 (36.46)	101 (27.7)	264 (72.3)	1.4 (1.1-1.9)	.02
No	636 (63.54)	134 (21.1)	502 (78.9)	1.0	
Bullied (last 6 months)					
Yes	265 (26.47)	91 (34.3)	174 (65.7)	2.2 (1.6-2.9)	<.001
No	736 (73.53)	144 (19.6)	592 (80.4)	1.0	
Misused detox/laxative teas or diet pills (last 12 months)					
Yes	83 (8.29)	48 (57.8)	35 (42.2)	5.3 (3.4-8.5)	<.001
No	918 (91.71)	187 (20.4)	731 (79.6)	1.0	
Ever used illegal drugs					
Yes	545 (54.45)	107 (19.6)	438 (80.4)	1.0	
No	446 (44.56)	126 (28.3)	320 (71.8)	1.6 (1.2-2.2)	.002

Variable	Total, n (%) N=1001	Consumers, n (%) n=235	Nonconsumers, n (%) n=766	OR (95% CI)	P
Last month illegal drug use if ever used drugs					
Yes	330 (33.00)	72 (21.8)	258 (78.2)	1.4 (0.9-2.2)	.11
No	215 (21.48)	35 (16.3)	180 (83.7)	1.0	
Risky single occasion drinking weekly or more often					
Yes	122 (12.19)	21 (17.2)	101 (82.8)	1.0	
No	758 (75.72)	179 (23.6)	579 (76.4)	1.5 (0.9-2.4)	.12
Current smoker					
Yes	206 (20.68)	47 (22.8)	159 (77.2)	1.0 (0.7-1.5)	.83
No	790 (79.32)	186 (23.5)	604 (46.5)	1.0	

[a] Gay, lesbian, bisexual, queer, or questioning.

[b] Based on self-reported diagnosed and undiagnosed conditions in the last 6 months.

Discussion

Our study, to the best of our knowledge, was the first to explore characteristics of the consumers of health and fitness-related social media content. Our results indicate that consuming health and fitness-related social media content is common; 378 of 1001 (37.76%) participants reported liking or following at least one of the included health and fitness-related social media content types on Facebook, Instagram, or Twitter, most commonly fitspiration pages (308/1001, 30.8%), followed by diet/fitness plan pages (235/1001, 23.5%), and detox pages (145/14.49, 14.5%). The majority of health and fitness-related social media content consumers identified as female, supporting our hypothesis. This result was unsurprising; health and fitness-related social media content is largely aimed at women and often driven by female celebrities and fitness models. Considering the number of objectifying messages previously observed in fitspiration [4], and the potential internalization of messages such as these by girls and women in Western society [23], it is potentially concerning that nearly half of female participants reported consuming this content. Even so, some health and fitness-related social media content is aimed at men (ie, bodybuilding pages featuring endorsement from male athletes) and there is potential for this content to negatively affect the body image of young men, such as increasing a drive for muscularity [24].

Other demographic differences were noted fairly consistently in the data. Key characteristics of health and fitness-related social media content consumers were being younger and less educated; more than half of participants aged between 15 and 17 years and more than half of participants with no post-high school education (which included those still in high school) reported consuming at least one of the health and fitness-related social media content types, although this latter variable was not significant in adjusted analyses. In all, nearly half of all consumers (48.7%, 184/378) were teenaged girls. These findings are of concern because adolescence is a particularly challenging time in terms of body image [25] and more educated people are generally more likely to engage in healthy behaviors, such as

engaging in physical activity and not smoking [26], and have higher health literacy [27].

Some differences were observed regarding mental health and substance use, partially supporting our hypothesis. Participants with eating disorders were 2 to 3 times more likely to consume health and fitness-related social media content than participants without eating disorders. It is likely that this relationship is bidirectional; this content may attract people with eating disorders or body image concerns, but the content may exacerbate or validate symptomology and behaviors [8]. Further, a significant difference emerged with regards to mood disorders: participants with mood disorders were more likely than those without mood disorders to consume diet/fitness plan pages, although this was not significant in adjusted analyses. This finding is interesting in the context of thinspiration research, which has found more negative affect after viewing thinspiration websites [14]. It is unclear why this relationship emerged for the diet/fitness plan pages, but not the other types of health and fitness-related social media content.

Approximately 70% of participants who reported misusing detox/laxative teas or diet pills in the last 6 months consumed any health and fitness-related social media content, supporting our hypothesis. These weight loss materials have been shown to have detrimental health effects and use actually predicts weight gain over time in adolescents [28,29]. Consumers of any health and fitness-related social media content were significantly less likely than nonconsumers to report ever using illegal drugs or to report weekly risky single occasion drinking. This was an interesting finding; it is possible that consumers of health and fitness-related social media content are concerned about the effect of substance use on their health, thereby avoiding consumption. However, several of these results were not significant in adjusted analyses and it is possible that these findings were related to the age of participants because younger participants were more likely to consume health and fitness-related social media content. In adjusted analyses, consumers of detox pages were more likely to have used illegal drugs recently than nonconsumers. This association may be related to use of illegal drugs for weight loss or maintenance,

such as psychostimulants [30]. These relationships should be explored in future research.

It is possible that for the majority of consumers, health and fitness-related social media content is beneficial and provides motivation for healthy behaviors. Yet the key characteristics of consumers of health and fitness-related social media content appear to be female gender and a younger age, with at-risk groups including those with eating disorder symptomology, being a victim of bullying, and misusing detox/laxative teas and diet pills. Health and fitness-related social media content potentially has the power to impact on these individuals, including influencing the formation of young people's norms regarding ideal body shape and what it means to be healthy; emerging research indicates that adolescent girls are increasingly turning to fitness models as role models [31]. It can be difficult to distinguish between health and fitness-related social media content that is helpful or motivational and content that is harmful, and to whom messages championed by this content are negatively affecting; however, it appears that some vulnerable individuals are consuming health and fitness-related social media content.

Nearly 90% of American young adults have reported they would trust medical information found on social media [19]. Therefore, there is a need to ensure that health and fitness-related social media content portrays adequate, responsible health messages championing accurate information about health and fitness, motivating individuals to exercise without shaming those who do not or cannot, having realistic health and fitness goals, and encouraging a healthy lifestyle without objectifying messages. Health promotion initiatives should target consumers of health and fitness-related social media content in terms of health literacy and body positivity, teaching at-risk individuals to be critical of media messages in relation to what it means to be fit and healthy. Some recent campaigns have attempted this, such as the UK campaign "This Girl Can," which aims to celebrate women's participation in sport regardless of physical appearance [32].

Another possible option for dealing with potential harms of health and fitness-related social media content is regulating social media content, although this can be challenging. Thinspiration is recognized as harmful by most social media sites and is shut down or censored with varying degrees of success [33]; for example, Facebook community standards state "We prohibit content that promotes or encourages...eating disorders" [34] and if users search for thinspiration and related terms, Instagram and Tumblr provide warnings for graphic content and referrals to eating disorder information and recovery resources. However, due to varying rates of effective moderation and social media sites not wishing to censor users' recovery journeys, it is still easy to find thinspiration content across nearly every social media platform [35]. No specific guidelines exist for health and fitness-related social media content. Current advertising guidelines on Facebook indicate that images that "emphasize an 'ideal' body or body parts, or images showing unexpected or unlikely results, such as 'before and after images'" are not allowed, and that "ads that promote acceptable dietary and herbal supplements may only target users who are at least 18 years of age" [35], but such guidelines do not appear

to exist for page or user-generated content, even when the page is advertising a product. Considering problematic health and fitness-related social media content messages may be subtle or labeled as "healthy," these guidelines may not adequately identify harmful content.

Two additional barriers to regulating health and fitness-related social media content on social media have been raised. Firstly, social media is saturated with health and fitness-related social media content: this content has huge followings (eg, more than 23 million posts on Instagram have been tagged with "#fitspo" at the time of writing) and the nature of social media means that health and fitness-related social media content is often viewed by social media users even if they do not necessarily wish to view it (ie, if one user "likes" an image, this image will then appear in the newsfeeds of many of their friends). The nature of targeted advertisements means that merely mentioning food or exercise on social media can result in users being presented with advertisements related to health and fitness-related social media content [5]. Secondly, health and fitness-related social media content is largely celebrated, user-generated, and talked about in a positive manner, reinforcing content and behaviors to peers on social media [3]. An argument for clinically distinguishing orthorexia from anorexia and obsessive-compulsive disorder is that people with orthorexia are likely to flaunt their health behaviors, such as via social media [21], because these behaviors are largely socially desirable and celebrated, making it difficult to determine where health behaviors are obsessive and/or no longer healthy. The saturation and popularity of health and fitness-related social media content means that its messages are unavoidable for many users of social media and easily normalized regardless of actual health benefits. This reinforces the importance of media literacy and education programs around health and fitness for young people.

The authors recognize the limitations of this study. The sample was an online convenience sample and may not be generalizable to all social media users. The questions asked were broad and lacked specificity; we did not enquire after the number of health and fitness-related social media content pages liked or followed, the degree of interaction with the content, or break down the pages any further (eg, by examining participants who specifically followed self-labeled fitspiration pages on social media). Data were self-reported and, therefore, vulnerable to recall bias; social media users often follow a large number of pages [36] and are unlikely to remember all of them. Those who consumed fitspiration pages, detox pages, and diet/fitness plan pages were more likely to follow other health pages too, possibly reflecting an interest in health in general rather than just the 3 types of health and fitness-related social media content we studied. We only focused on Facebook, Instagram, or Twitter, potentially excluding participants who follow health and fitness-related social media content on Tumblr or Pinterest (used by 23% and 33% of teenage girls, respectively, more than 3 times the rate of use by teenaged boys [2]) or engage with other user-generated forums and groups such as those on Reddit. We only asked about participants misusing diet pills and detox/laxative teas; it would have been worth exploring any use of diet/weight loss materials. Our cross-sectional design and

analysis strategy was unable to determine direction of relationships or causality. Due to the length of the larger survey, we did not examine body image or use validated measures of mental health, although single-item self-report measures of psychosocial variables can be as valid as multiple-item scales [37]. Further, due to the exploratory nature of the study, we included a large number of statistical tests; we applied the conservative Bonferroni correction, which thereby increases the risk of type II error.

Observed gender differences are likely related to the types of health and fitness–related social media content we chose to examine. The diet plans and challenges included as examples were heavily female-led (eg, branded with female celebrities) and focused (eg, bikini body challenges), potentially biasing recall. Gender differences may have also been related to women consuming more health-related pages on social media than men in general. We did not specifically examine, or include as examples, health and fitness–related social media content aimed at men, such as bodybuilding or other muscularity-based initiatives. Such health and fitness–related social media content is worth exploring in the future.

This is the first study to characterize consumers of 3 types of health and fitness-related social media content: fitspiration pages, detox pages, and diet/fitness plan pages. Overall, the results of this exploratory study indicate that the consumers of health and fitness-related social media content are largely teenaged girls and that individuals reporting eating disorders and detox or laxative misuse are more likely to consume health and fitness-related social media content. The results emphasize the need to perform further research into this area and consider the role of health and fitness-related social media content in the formation of body image, health ideals and behaviors, and emerging mental health issues such as orthorexia within the complex context of normative processes and development, particularly among at-risk individuals [8]. Future experimental or longitudinal research should determine whether health and fitness-related social media content actually impacts the consumer's body image and health behaviors and, if so, how it can be addressed. There is also a need to perform a content analysis on health and fitness-related social media content to determine to what degree these pages are championing accurate versus unhealthy or unscientific health messages and determine which social media platforms are best to target for future interventions.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Example of fitspiration page: 'Fitspore' profile on Instagram.

[PNG File, 1MB - jmir_v17i8e205_app1.png]

Multimedia Appendix 2

Example of detox page: 'SkinnyMe Tea' on Facebook.

[PNG File, 529KB - jmir_v17i8e205_app2.png]

Multimedia Appendix 3

Example of diet/fitness plan page: 'Ashy Bines Bikini Body Challenge' on Facebook.

[PNG File, 969KB - jmir_v17i8e205_app3.png]

References

1. Duggan M, Ellison N, Lampe C, Lenhart A, Madden M. Social Media Update 2014. Washington, DC: Pew Internet & American Life Project; 2015 Jan 09. URL: http://www.pewinternet.org/files/2015/01/PI_SocialMediaUpdate20144.pdf [accessed 2015-08-06] [WebCite Cache ID 6aZDcWwMQ]
2. Lenhart A. Teens, Social Media & Technology Overview 2015. Washington, DC: Pew Internet & American Life Project; 2015 Apr 09. URL: http://www.pewinternet.org/files/2015/04/PI_TeensandTech_Update2015_0409151.pdf [accessed 2015-08-06] [WebCite Cache ID 6aZDhce1H]
3. Vaterlaus J, Patten E, Roche C, Young J. #Gettinghealthy: The perceived influence of social media on young adult health behaviors. Computers in Human Behavior 2015 Apr;45:151-157. [doi: 10.1016/j.chb.2014.12.013]

XSL•FO

RenderX

4. Boepple L, Thompson JK. A content analytic comparison of fitspiration and thinspiration websites. Int J Eat Disord 2015 Mar 16:epub. [doi: 10.1002/eat.22403] [Medline: 25778714]

5. Villiard H, Moreno MA. Fitness on Facebook: advertisements generated in response to profile content. Cyberpsychol Behav Soc Netw 2012 Oct;15(10):564-568 [FREE Full text] [doi: 10.1089/cyber.2011.0642] [Medline: 22963337]

6. Mabe AG, Forney KJ, Keel PK. Do you "like" my photo? Facebook use maintains eating disorder risk. Int J Eat Disord 2014 Jul;47(5):516-523. [Medline: 25035882]

7. Fardouly J, Diedrichs PC, Vartanian LR, Halliwell E. Social comparisons on social media: the impact of Facebook on young women's body image concerns and mood. Body Image 2015 Mar;13:38-45. [doi: 10.1016/j.bodyim.2014.12.002] [Medline: 25615425]

8. Perloff RM. Social media effects on young women's body image concerns: theoretical perspectives and an agenda for research. Sex Roles 2014 May 29;71(11-12):363-377. [doi: 10.1007/s11199-014-0384-6]

9. Tate DF, Lyons EJ, Valle CG. High-tech tools for exercise motivation: use and role of technologies such as the internet, mobile applications, social media, and video games. Diabetes Spectr 2015 Jan;28(1):45-54. [doi: 10.2337/diaspect.28.1.45] [Medline: 25717278]

10. Markula P. Firm but shapely, fit but sexy, strong but thin: the postmodern aerobicizing female bodies. Sociology of Sport Journal 1995;12(4):424-453.

11. Homan K. Athletic-ideal and thin-ideal internalization as prospective predictors of body dissatisfaction, dieting, and compulsive exercise. Body Image 2010 Jun;7(3):240-245. [doi: 10.1016/j.bodyim.2010.02.004] [Medline: 20226748]

12. Kite L. Huffington Post Women. 2015 May 17. 'Fitspiration': Why it isn't so inspirational URL: http://www.huffingtonpost.com/lexie-kite/fitspiration-isnt-inspirational_b_1524706.html?ir=Women&ref=fb&src=sp&comm_ref=false [WebCite Cache ID 6Z1hzfOBr]

13. Low KG, Charanasomboon S, Brown C, Hiltunen G, Long K, Reinhalter K, et al. Internalization of the thin ideal, weight, and body image concerns. Soc Behav Pers 2003 Jan 01;31(1):81-89. [doi: 10.2224/sbp.2003.31.1.81]

14. Bardone-Cone A, Cass K. Investigating the impact of pro-anorexia websites: a pilot study. Eur Eat Disorders Rev 2006 Jul;14(4):256-262. [doi: 10.1002/erv.714]

15. Jett S, LaPorte DJ, Wanchisn J. Impact of exposure to pro-eating disorder websites on eating behaviour in college women. Eur Eat Disord Rev 2010;18(5):410-416. [doi: 10.1002/erv.1009] [Medline: 20572210]

16. Wilson JL, Peebles R, Hardy KK, Litt IF. Surfing for thinness: a pilot study of pro-eating disorder web site usage in adolescents with eating disorders. Pediatrics 2006 Dec;118(6):e1635-e1643. [doi: 10.1542/peds.2006-1133] [Medline: 17142493]

17. British Dietetic Association. 2015 Dec 08. Top 5 worst celebrity diets to avoid in 2015 URL: https://www.bda.uk.com/news/view?id=39&x[0]=news/list [WebCite Cache ID 6Z1iBW0sp]

18. NewsMail. 2014 Jan 07. Dietitians reveal the three worst diets for your health URL: http://www.news-mail.com.au/news/dietitians-reveal-three-worst-diets-your-health/2132566/ [accessed 2015-06-04] [WebCite Cache ID 6Z1iNbcZ9]

19. PwC Health Research Institute. Social Media "Likes" Healthcare: From Marketing to Social Business. 2012 Apr. URL: http://download.pwc.com/ie/pubs/2012_social_media_likes_healthcare.pdf [accessed 2015-08-06] [WebCite Cache ID 6aZDsazY0]

20. Boepple L, Thompson JK. A content analysis of healthy living blogs: evidence of content thematically consistent with dysfunctional eating attitudes and behaviors. Int J Eat Disord 2014 May;47(4):362-367. [doi: 10.1002/eat.22244] [Medline: 24420676]

21. Koven NS, Abry AW. The clinical basis of orthorexia nervosa: emerging perspectives. Neuropsychiatr Dis Treat 2015;11:385-394 [FREE Full text] [doi: 10.2147/NDT.S61665] [Medline: 25733839]

22. Babor TF, Higgins-Biddle JC, Saunders JB, Monteiro MG. AUDIT: the Alcohol Use Disorders Identification Test. Guidelines for Use in Primary Care. 2nd ed. Geneva: World Health Organization; 2001. URL: http://apps.who.int/iris/bitstream/10665/67205/1/WHO_MSD_MSB_01.6a.pdf [accessed 2015-08-16] [WebCite Cache ID 6apzSzAT0]

23. Fredrickson BL, Roberts TA. Objectification theory: toward understanding women's lived experiences and mental health risks. Psychology of Women Quarterly 1997;21(2):173-206. [doi: 10.1111/j.1471-6402.1997.tb00108.x]

24. Pritchard M, Cramblitt B. Media influence on drive for thinness and drive for muscularity. Sex Roles 2014 Aug 2;71(5-8):208-218. [doi: 10.1007/s11199-014-0397-1]

25. Kostanski M, Gullone E. Adolescent body image dissatisfaction: relationships with self-esteem, anxiety, and depression controlling for body mass. J Child Psychol Psychiatry 1998 Feb;39(2):255-262. [Medline: 9669238]

26. Pampel FC, Krueger PM, Denney JT. Socioeconomic disparities in health behaviors. Annu Rev Sociol 2010 Aug;36:349-370 [FREE Full text] [doi: 10.1146/annurev.soc.012809.102529] [Medline: 21909182]

27. Australian Bureau of Statistics. 2009 Jun 30. 4102.0-Australian Social Trends, June 2009 URL: http://www.abs.gov.au/AUSSTATS/abs@.nsf/Lookup/4102.0Main+Features20June+2009 [accessed 2015-06-04] [WebCite Cache ID 6Z1igg0s1]

28. Neumark-Sztainer D, Wall M, Story M, Standish AR. Dieting and unhealthy weight control behaviors during adolescence: associations with 10-year changes in body mass index. J Adolesc Health 2012 Jan;50(1):80-86 [FREE Full text] [doi: 10.1016/j.jadohealth.2011.05.010] [Medline: 22188838]

29.   Yen M, Ewald MB. Toxicity of weight loss agents. J Med Toxicol 2012 Jun;8(2):145-152 [FREE Full text] [doi: 10.1007/s13181-012-0213-7] [Medline: 22351299]

30.   Neale A, Abraham S, Russell J. "Ice" use and eating disorders: a report of three cases. Int J Eat Disord 2009 Mar;42(2):188-191. [doi: 10.1002/eat.20587] [Medline: 18803173]

31.   Yasa D. Body and Soul. 2015. Can you trust an Instagram expert? URL: http://www.bodyandsoul.com.au/health/health+news/can+you+trust+an+instagram+expertr,36777 [accessed 2015-06-04] [WebCite Cache ID 6Z1j4Z119]

32.   Gearin M. ABC News. 2015 Feb 21. This Girl Can advertising campaign encouraging women and teenage girls to play sport goes viral URL: http://www.abc.net.au/news/2015-02-22/this-girl-can-viral-campaign-encouraging-females-in-sport/6175398 [accessed 2015-06-04] [WebCite Cache ID 6Z1jBvhSH]

33.   Spiegel A. Buzzfeed News. 2013 Jan 11. Here's how social media 'Thinspiration' bans are actually working out URL: http://www.buzzfeed.com/verymuchso/heres-how-social-media-thinspiration-bans-are-a [accessed 2015-06-04] [WebCite Cache ID 6Z1jGYIFB]

34.   Facebook. Self-injury community standards URL: https://www.facebook.com/communitystandards/ [accessed 2015-08-16] [WebCite Cache ID 6aq0FAdEh]

35.   Facebook. Prohibited content URL: https://www.facebook.com/policies/ads/#prohibited_content [accessed 2015-08-17] [WebCite Cache ID 6aq0IBwcc]

36.   Socialbakers. 2014. Cutting through the crowds on Facebook news feeds URL: http://www.socialbakers.com/blog/1561-cutting-through-the-crowds-on-facebook-news-feeds [accessed 2015-06-04] [WebCite Cache ID 6Z1jR4TK4]

37.   Hoeppner BB, Kelly JF, Urbanoski KA, Slaymaker V. Comparative utility of a single-item versus multiple-item measure of self-efficacy in predicting relapse among young adults. J Subst Abuse Treat 2011 Oct;41(3):305-312 [FREE Full text] [doi: 10.1016/j.jsat.2011.04.005] [Medline: 21700411]

Abbreviations

GLBQQ+: gay, lesbian, bisexual, queer, or questioning

Original Paper

Analyzing Information Seeking and Drug-Safety Alert Response by Health Care Professionals as New Methods for Surveillance

Alison Callahan[1], BSc (Hons), MISt, PhD; Igor Pernek[2], PhD; Gregor Stiglic[3,4], PhD; Jure Leskovec[5], PhD; Howard R Strasberg[6], MD, MS; Nigam Haresh Shah[1], MBBS, PhD

[1]Stanford Center for Biomedical Informatics Research, Stanford University, Stanford, CA, United States

[2]Pervasive Computing Applications, Research Studios Austria, Vienna, Austria

[3]Faculty of Health Sciences, University of Maribor, Maribor, Slovenia

[4]Faculty of Electrical Engineering and Computer Science, University of Maribor, Maribor, Slovenia

[5]Computer Science Department, Stanford University, Stanford, CA, United States

[6]Wolters Kluwer Health, San Diego, CA, United States

Corresponding Author:
Alison Callahan, BSc (Hons), MISt, PhD
Stanford Center for Biomedical Informatics Research
Stanford University
Room X-215
1265 Welch Road
Stanford, CA, 94305-5479
United States
Phone: 1 6507236979
Fax: 1 650 725 7944
Email: acallaha@stanford.edu

Abstract

Background: Patterns in general consumer online search logs have been used to monitor health conditions and to predict health-related activities, but the multiple contexts within which consumers perform online searches make significant associations difficult to interpret. Physician information-seeking behavior has typically been analyzed through survey-based approaches and literature reviews. Activity logs from health care professionals using online medical information resources are thus a valuable yet relatively untapped resource for large-scale medical surveillance.

Objective: To analyze health care professionals' information-seeking behavior and assess the feasibility of measuring drug-safety alert response from the usage logs of an online medical information resource.

Methods: Using two years (2011-2012) of usage logs from UpToDate, we measured the volume of searches related to medical conditions with significant burden in the United States, as well as the seasonal distribution of those searches. We quantified the relationship between searches and resulting page views. Using a large collection of online mainstream media articles and Web log posts we also characterized the uptake of a Food and Drug Administration (FDA) alert via changes in UpToDate search activity compared with general online media activity related to the subject of the alert.

Results: Diseases and symptoms dominate UpToDate searches. Some searches result in page views of only short duration, while others consistently result in longer-than-average page views. The response to an FDA alert for Celexa, characterized by a change in UpToDate search activity, differed considerably from general online media activity. Changes in search activity appeared later and persisted longer in UpToDate logs. The volume of searches and page view durations related to Celexa before the alert also differed from those after the alert.

Conclusions: Understanding the information-seeking behavior associated with online evidence sources can offer insight into the information needs of health professionals and enable large-scale medical surveillance. Our Web log mining approach has the potential to monitor responses to FDA alerts at a national level. Our findings can also inform the design and content of evidence-based medical information resources such as UpToDate.

(J Med Internet Res 2015;17(8):e204)   doi:10.2196/jmir.4427

XSL•FO
RenderX

Introduction

Searching and consuming medical information resources on the Web occupies an increasingly important place in both consumer and health care professionals' day-to-day information-seeking activities [1-6]. Methods for mining Web search logs to characterize user behavior and to perform large-scale surveillance—such as Google Flu Trends [7] and similar efforts using Wikipedia Web traffic [8]—are gaining traction. Studies have sought to characterize the search behavior of specific user groups, for example those who search for information about cancer [9] or varicose vein treatment [10], as well as the effect of significant large-scale events like the recession [11] or the time of year [12] on searches related to health concerns. Patterns in search logs have also been used to predict health-related activities, including visits to medical facilities [13] and the onset of searches about mood-stabilizing drugs [14], as well as to track changes in drug use over time [15].

A challenge inherent in the analysis of Web search behavior generally is the diversity of users. Web search logs capture a broad range of online behavior from a largely uncharacterized user group performing searches with unknown context(s). Most efforts at analyzing Web logs focus on consumer search behavior. In fact, recent methods attempt to discern and separate out searches from health care professionals in the analysis [16].

In contrast, the analysis of search behavior of health care professionals has typically focused on literature review or survey-based approaches [4,17-20]. To directly study information seeking by health care professionals "in the wild" we present an analysis of activity logs from a widely used online medical resource called UpToDate [21]. UpToDate is a source of expert-authored health information provided by Wolters Kluwer that includes detailed descriptions of approaches to investigating specific symptoms, management of diseases, drug usage recommendations, and treatments to support evidence-based medicine. UpToDate is used on a subscription basis by institutions and individuals who purchase a license, including physicians, researchers, and students. Its use in hospitals is known to be associated with fewer patient complications and adverse events, shorter hospital stays, reduced mortality rates, and higher quality performance measures [22,23]. Given the challenges in analyzing general Web search behavior, the UpToDate logs are a unique resource—they capture the search behavior of a limited and well-defined user group. UpToDate logs have been previously used to predict flu trends in a timely manner [24], demonstrating the utility of this resource as an alternative data source for medical surveillance.

Logs of UpToDate usage capture the source institution and a unique deidentified session, the search string entered, the time and date of the search, the type of search, and topic pages visited as a result of a search. Using these access logs, we analyzed both the free-text searches users performed as well as how they navigated UpToDate topic pages. We profiled how UpToDate is used nationally and we quantified the relationship of medical conditions to cost and utilization as seen from an information-seeking perspective. We also characterized information-seeking behavior via the relationship between search terms and subsequent page view duration—a well-studied indicator of user interest [25,26]. We identified patterns in UpToDate page view sequences and the search terms that initiated them. We also present results on the use of UpToDate logs to monitor uptake of a Food and Drug Administration (FDA) alert.

In the following sections, we describe our methods for analyzing the free-text searches and page view sequences from UpToDate search logs. We then present findings of seasonal distributions of UpToDate free-text searches, as well as the distribution of searches by body systems, health conditions and symptoms, drugs, medical devices, and procedures. We also describe the uptake of an FDA drug alert as reflected by changes in the frequency and duration of UpToDate searches and page views. Lastly, we discuss the implications of our findings, describe the limitations of our approach, and propose future work.

Methods

Overview

To profile the online search behavior of health care professionals, we used 2 years of usage logs from UpToDate, spanning from January 2011 to December 2012. We analyzed these logs using a combination of text mining and statistical methods to identify general trends in searches, to discover associations between search terms and the duration of subsequent visits to UpToDate topic pages ("topic views") as well as patterns in topic view sequences, and to monitor the uptake of an FDA alert by health care professionals. In the following sections, we describe the structure of the search logs, our text-mining approach, and methods for analyzing UpToDate user behavior as well as identifying patterns in searches and topic views.

Structure of UpToDate Search Logs

The log for a single UpToDate user event consists of the following parts: (1) query string, (2) unique session ID, (3) search location, (4) time stamp of the search, and (5) the type of action (eg, a string search in the search bar of the website, a topic view, or a subtopic view, which is recorded when a user clicks on a link within an UpToDate page).

We restricted this dataset to consider only searches or page views performed at sites in the United States that purchased an UpToDate license (ie, we have excluded users of trial or marketing versions of UpToDate, as well as those searches performed on computers outside of the United States). The dataset used contains 212 million search queries and their corresponding topic views.

Text Processing of Search Logs

We processed all free-text searches using a variation of our previously described text-processing workflow [27,28]. We

used the Unitex corpus processor to annotate search strings with a lexicon of more than 3 million terms compiled from biomedical ontologies and terminologies, in which terms and concepts are mapped by synonymy and parent-child relationships. The output of this annotation process is a list of terms for each query string, each of which is mapped to one or more biomedical concepts. Concepts in the lexicon are in turn mapped to one of four semantic types, where applicable—diseases and symptoms, drugs, medical devices, and procedures.

Analyzing Free-Text Searches

Using the output of our text-processing pipeline, we computed the frequency of occurrence of each term across all usage logs, and used these frequencies to profile how health care professionals search UpToDate. We first profiled search behavior using the following features: seasonality, search session length, body system, medical concept category, and national statistics on the costs and discharge rates of medical conditions in 2011 and 2012. We used the time stamp associated with each search to aggregate searches by month and year. We used the concept mappings in our terminology to identify term sets for the 10 major body systems: cardiovascular, respiratory, digestive, endocrine, hemic and immune, integumentary,

musculoskeletal, mouth and jaw, nervous, and urogenital. Using these term sets, we counted the number of searches containing any of the terms associated with a given body system. Using the concept mappings in our terminology, we determined the most frequently searched-for disease/symptom, drug, device, and procedure terms, and quantified the distribution of these categories for each body system to assess the comprehensiveness of search logs as a data source.

To explore relationships between the search behavior and national trends of medical conditions in the US, we created custom term sets from our lexicon corresponding to the Healthcare Cost and Utilization Project (HCUP) clinical classification system (CCS) codes. We started with the International Classification of Diseases, 9th Revision (ICD-9) codes that make up each CCS code, and collected the Unified Medical Language System (UMLS) Metathesaurus concepts for each of the ICD-9 codes. We then extended the terms sets associated with these concepts by leveraging the term-concept mappings in our much larger lexicon (see Figure 1), and manually reviewed the terms sets to remove overly broad or incorrect terms. We used these curated term sets to identify searches in UpToDate that are related to medical conditions or procedures for which national data are available from the HCUP National Inpatient Sample.

Figure 1. Example of term expansion for HCUP CCS codes; each code is represented as its set of ICD-9 codes. The ICD-9 code for essential hypertension is expanded using the UMLS Metathesaurus to identify the n terms mapped to that concept. Each of the n terms (eg, "essential hypertension nos," red) is used as a seed query to our custom lexicon to identify additional concepts and their terms. Here, "essential hypertension nos" maps to the concept hypertensive disease, and its additional m-1 terms (green) are used in combination with the n seed terms to identify searches corresponding to the HCUP CCS codes for hypertension.
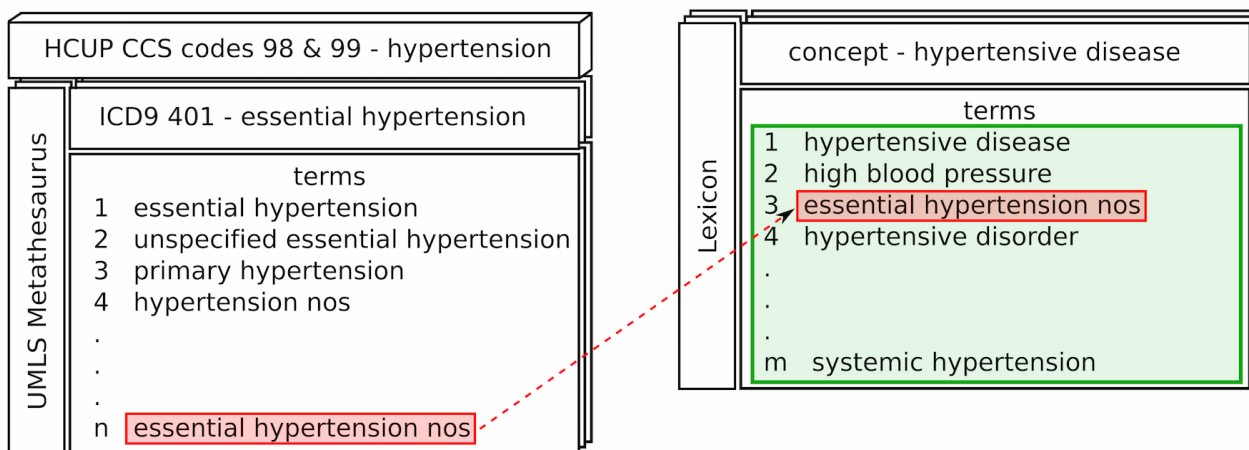
Analyzing Information-Seeking Behavior

We used time stamps for all searches and topic views to determine duration of topic views immediately following a given search. Topic view duration was calculated by taking the difference between the time stamp of a topic view event and the time stamp of the next event that followed it—either a search or view of a completely different topic. If the final event within a session was a topic view, that topic view event was excluded from this part of our analysis because it was not possible to calculate the duration of this topic view without the logged time stamp of a subsequent user action. From the log-normalized distribution of time spent on all topic pages following a search, we categorized the amount of time spent on a given topic page

using the mean of the log-normalized duration values (converted back to seconds) as the decision boundary. Durations less than this boundary were categorized as *short clicks*, and durations above this boundary were categorized as *long clicks*.

We grouped all searches that initiated a given topic view to calculate the proportion of long clicks and short clicks for each topic, and to determine the number of unique search terms resulting in a topic view. We also grouped all topics resulting from a unique search term to calculate the proportion of long clicks and short clicks originating from the search term and the number of topics each search term initiated.

Mining Patterns of Topic View Sequences

UpToDate content is structured as topics, each of which has a dedicated page with subsections containing more specific information. In addition to analyzing the search text, we used the logs to study how UpToDate users progress from topic to topic within a session, and to profile the search terms that initiate sequences of topic views. Figure 2 summarizes our method. We first grouped searches by their unique session identifier and ordered the entries by the time stamp. For each sequence of two or more topic views, we counted the number of times that sequence occurred, as well as the search terms that occurred just prior to the sequence and their frequency.

Figure 2. Overview of the method used to analyze sequences of queries and topic views in UpToDate. Using unique session identifiers, we grouped queries and topic views, and ordered them by time of occurrence (green table). Frequently occurring sequences of topic views (orange boxes) were identified, as well as the search terms that initiated them (blue boxes). The numbers in parentheses show the frequency of occurrence of the specific term.

Mining UpToDate Logs to Measure Food and Drug Administration Alert Uptake

On August 24, 2011, the FDA published an alert describing the risk of adverse cardiovascular events associated with higher doses of citalopram (trade name Celexa). We examined the uptake of this alert in the UpToDate logs relative to consumer online media activity from the same time period. We used a large online media collection to compare the relative frequency of terms related to Celexa in the UpToDate search logs and in online news media. The online media collection consisted of over 6 billion online news articles, news wires, and blog posts published on the Web between 2009 and 2014 [29,30]. A single entry included the title, time stamp, URL of the article, as well as the article content. To obtain the documents, we used Spinn3r Web service [31] that monitors over 20 million Internet sources to retrieve approximately 3.2 million new documents each day. The collection represents a near-complete picture of US online media space.

We obtained the daily counts of the mentions of Celexa-related search terms in the news media collection and a cumulative daily count of all Celexa-related search queries on a specific day from the UpToDate logs. Using these counts, we calculated a 7-day moving average of the fraction of total daily search queries that were Celexa related. Similarly, we calculated daily counts and a 7-day moving average of the number of online media articles that contained Celexa-related terms. To enable an overlay of the UpToDate and online news counts, we scaled the UpToDate counts by 10^7. We considered a deviation from the mean number of occurrences over the 2-year time period as a signal.

Results

Overview

UpToDate use was dominated by searches for disease conditions and symptoms and showed significant seasonal variation. The conditions and symptoms that had a high query volume were

not those with the greatest burden on the US health care system. For example, headache and viral infections were in the top 20 medical conditions searched for in the study period, but have low relative aggregate cost and discharge rates. We also found that several medical conditions and procedures with high overall expenditures had low query volumes, such as heart attack and mood disorders. The top 1% of institutions by query volume were responsible for about 21% of queries in the United States, and issued an average of 1.7 million queries each. The rest of the queries were widely distributed nationally with some queries coming from each of the fifty states.

We summarized searches and the topic views that followed them in three ways. For topics, we found that topics with longer-than-average dwell times (also known as *long clicks*) had significantly fewer initiating search terms than those with shorter-than-average dwell times (*short clicks*). For search terms, we found that topic views initiated by a given search term were either, on average, longer than the mean topic view duration (ie, that search term always results in long clicks) or were, on average, shorter than the mean topic view duration (ie, that search term always results in short clicks). Finally, we elucidated patterns of frequently occurring search and topic view sequences, which typically started with a search for a disease term and ended with topics related to therapies for that disease.

We then characterized the "response" of the health professionals to a 2011 FDA drug alert—about the risk of abnormal heart rhythms when taking high doses of citalopram (Celexa)—in terms of the change in information-seeking behavior. Our comparison of the volume of searches about citalopram (Celexa) in UpToDate versus mentions of these terms in news sites following a 2011 FDA alert found that the uptake of alerts by health care professionals in their daily work differed markedly from that of the general public. Searches in UpToDate related to citalopram peaked more than 10 days after the FDA alert and the initial spike in mentions at news sites, but persisted for much longer.

Seasonal and Topic Trends in UpToDate User Search Behavior

Seasonal Distribution of UpToDate Searches

We measured the monthly frequency of the 10 most searched diseases and drugs in any month of 2011 and 2012 (see Figure 3). There were similar spikes in searches for influenza and Tamiflu (the trade name for oseltamivir, a flu medication) in the early winter of 2011, and winter to early spring of 2012. Searches for pneumonia also peaked in the winter months of both years. There was a sharp increase in the number of searches for diabetes insipidus in November of 2012, representing the disease with the highest search volume within a single month. The most searched-for drugs were vancomycin and Bactrim (the trade name for trimethoprim/sulfamethoxazole), antibiotics that are used to treat a variety of bacterial infections.

Figure 3. Most frequently searched diseases (top) and drugs (bottom) in 2011-2012.
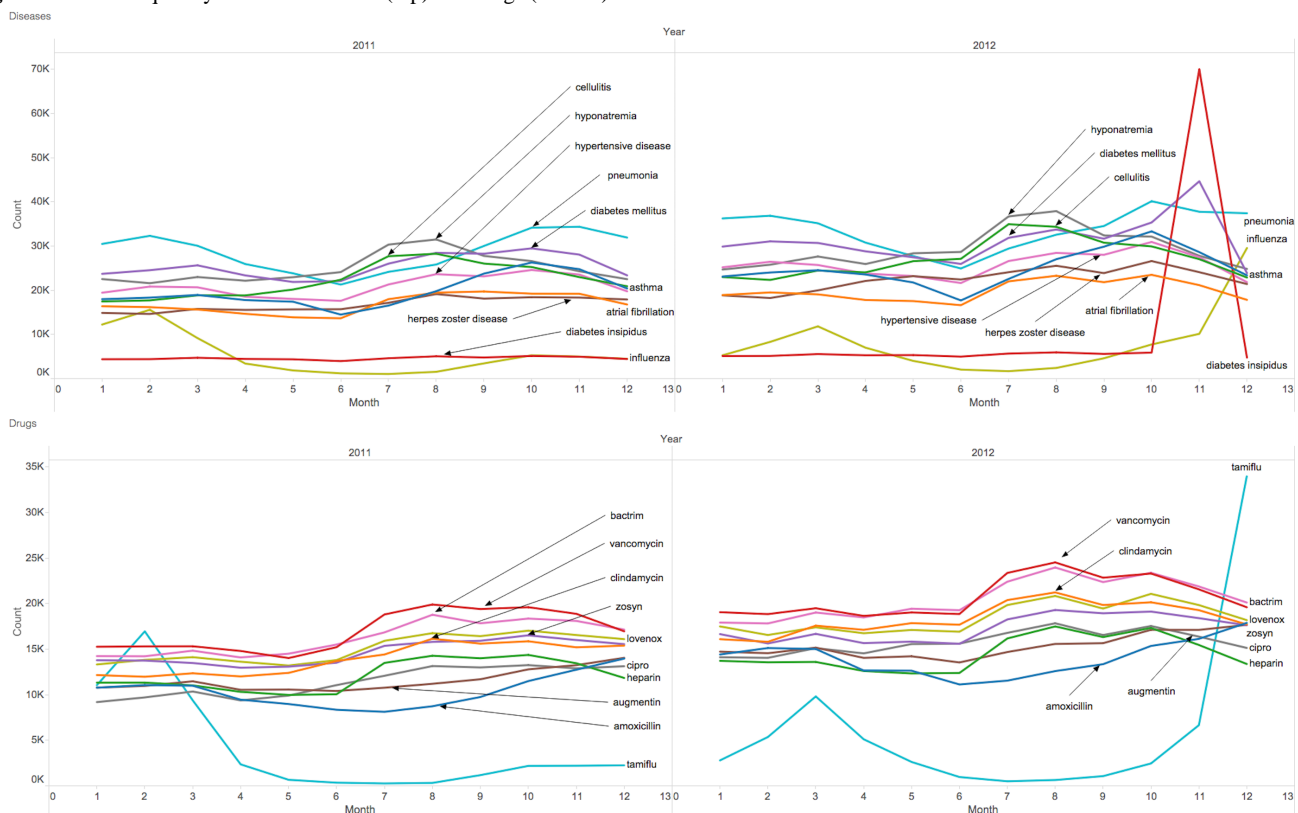
Distribution of Searches Across Organ Systems

Figure 4 shows the distribution by type of queries about each of the 10 major organ systems. The digestive system had the highest percentage of searches, followed closely by the cardiovascular system, while the mouth and jaw system had the lowest percentage of searches. Queries about all body systems were dominated by searches for diseases and symptoms. Searches about the endocrine system included a greater percentage of drugs. In general, queries related to medical

procedures were much lower in percentage, and were most represented in queries about the hemic and immune system, digestive system, and urogenital system. Figure 4 also lists the 10 most frequently and least frequently occurring terms related to each organ system (note that the categories are not exclusive,

because the same term may be related to multiple systems). Characterizing the relative search volume related to the major organ systems is important for assessing the feasibility of using the UpToDate search logs for large-scale surveillance.

Figure 4. Distribution of searches across organ systems. Each organ system name is followed by the percentage of queries related to that system. Each pie chart shows the distribution of searches for that organ system grouped by their term category, followed by the 10 most frequent and 10 least frequent terms searched for related to that system. Searches about diseases and symptoms (dark blue) dominated most systems. Searches about the endocrine system included a significant number of drug searches, followed by the cardiovascular system.
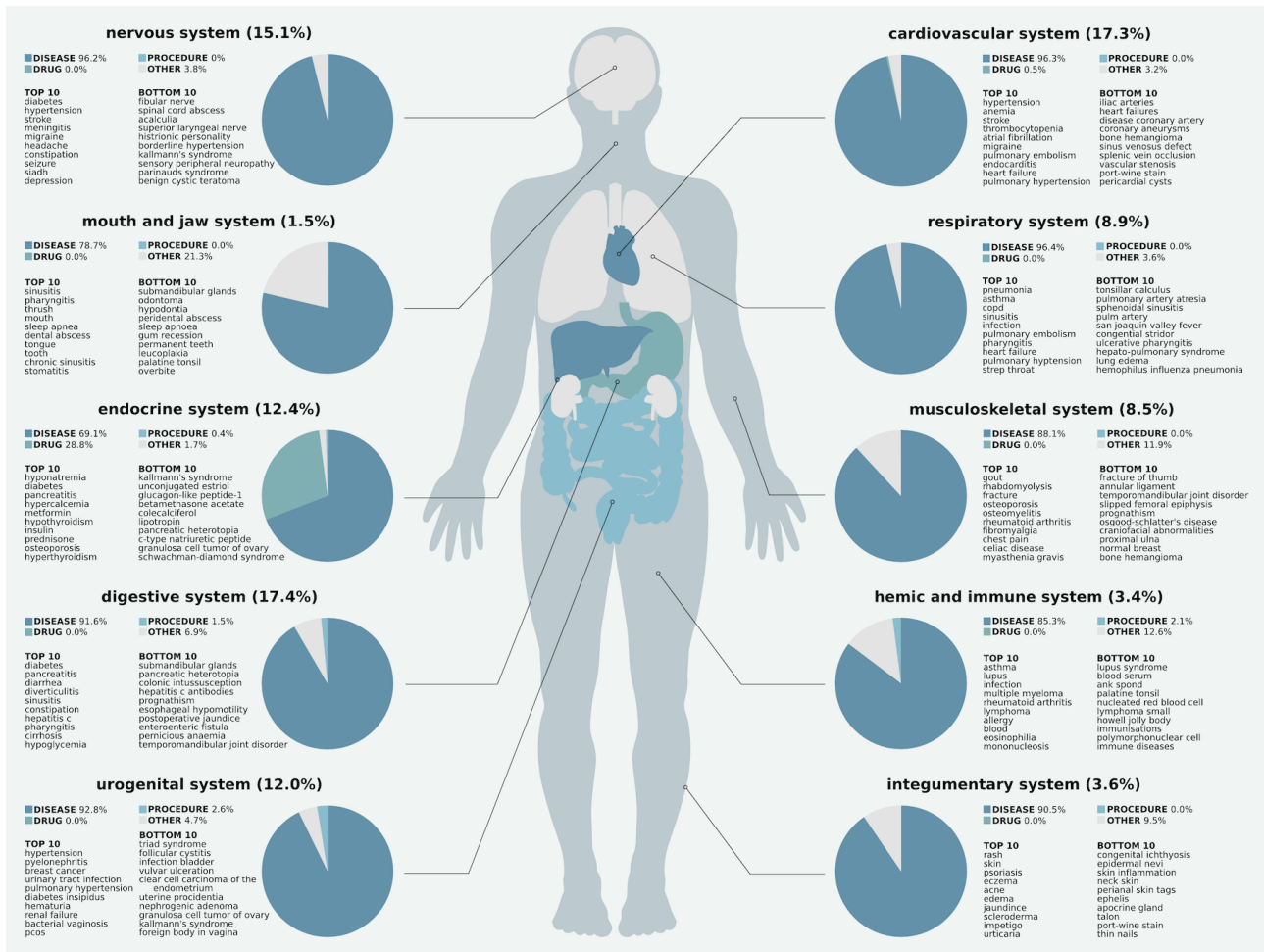
Distribution of Searches Across Medical Conditions

The query volume for the medical conditions with the most significant burden on the US health care system (quantified as being ranked within the top 10 mean aggregate costs and/or discharge volumes in 2011-2012) varied significantly across conditions (see Figure 5). Some of these medical conditions had very low query volumes in UpToDate—for example, live births, heart attack, and mood disorders—while others, such as septicemia and pneumonia, were frequently searched for. It is not surprising that conditions such as live birth or heart attack have low query volumes—these are medical events that have

well-understood management protocols, and thus it is expected that there is relatively little information need associated with them.

The most-searched medical conditions (the triangles in Figure 5) ranged from low aggregate cost and low discharge volume to high aggregate cost. Inflammatory skin conditions and infections had high query volume, but low aggregate cost and number of discharges. Septicemia and pneumonia were high on all three axes of cost, discharge, and query volume. These conditions with high query volume and burden represent surveillance opportunities for both public health and drug safety.

Figure 5. Medical conditions with the highest mean aggregate cost, discharge volume, and/or UpToDate query volume over 2011 and 2012. Each point is a single medical condition and the size of the point represents the query volume. The shape of the point indicates if it was ranked in the top 10 by aggregate cost, discharge volume, and/or query volume. Cost and discharge data were taken from the HCUP National Inpatient Sample.
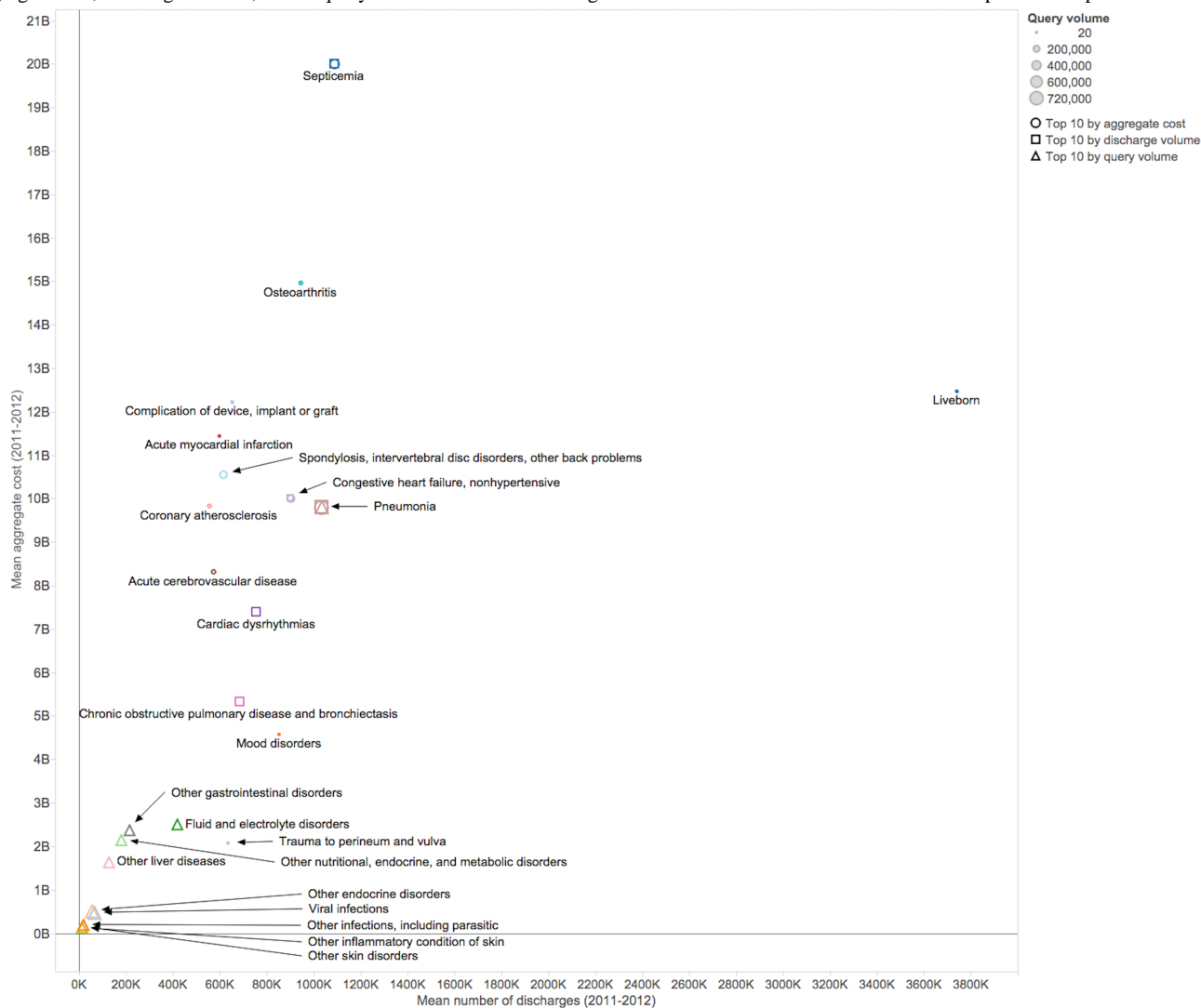
UpToDate User Behavior

To characterize how users search and consume UpToDate content, we analyzed the relationship between searches that users performed and the amount of time spent looking at the UpToDate topics returned as results for those searches. As described in the Methods section, we used the distribution of topic view durations to decide the cutoff duration (143.79 seconds) to categorize each topic view as a *short click* or a *long click*. We then determined the proportion of long clicks for each topic. The left panel of Figure 6 shows the distribution of long-click proportion across topics. We used the Hampel identifier for outliers to determine the threshold for high (0.70) and low (0.19) proportions of long clicks (shown as dotted lines in the left panel of Figure 6).

Topics with a high proportion of long clicks originated from significantly fewer unique searches than topics with mostly short clicks. Topics with a high proportion of long clicks had an average of 10.37 unique initiating search terms (SD 22.13) and a median of 4 (interquartile range [IQR] 2) unique initiating search terms. This is significantly lower than the average of

29.29 (SD 50.22) and median of 12 (IQR 8) unique initiating search terms for topics with a high proportion of short clicks (*P*<.001; Mann-Whitney U test).

In contrast, the distribution of long-click proportion for searches (the proportion of topic views with a long click initiated by a given search term) was distinctly bimodal (right panel of Figure 6). Searches tended to have either no long clicks following them, or mostly long clicks following them. This indicates that while topics typically had a number of search terms that landed the user on that topic and kept them reading for a long period of time, a given search term either always resulted in long clicks or always resulted in short clicks. Table 1 lists the top 10 search terms (by frequency in the logs) that always resulted in short clicks. Search terms that always result in short clicks may be terms that are too specific to return useful content or contain a typographical error (eg, "phenochromocytoma" and "probencid" are misspellings of the disease pheochromocytoma and the drug probenecid, respectively), but some may be candidates for addition to UpToDate content. Table 2 lists the top 10 search terms (by frequency in the logs) that always resulted in long clicks.

Figure 6. Distribution of long clicks in UpToDate topics (left) and search terms (right). The dotted lines in the left panel indicate the threshold for low and high long-click proportions determined using the Hampel identifier for outliers.
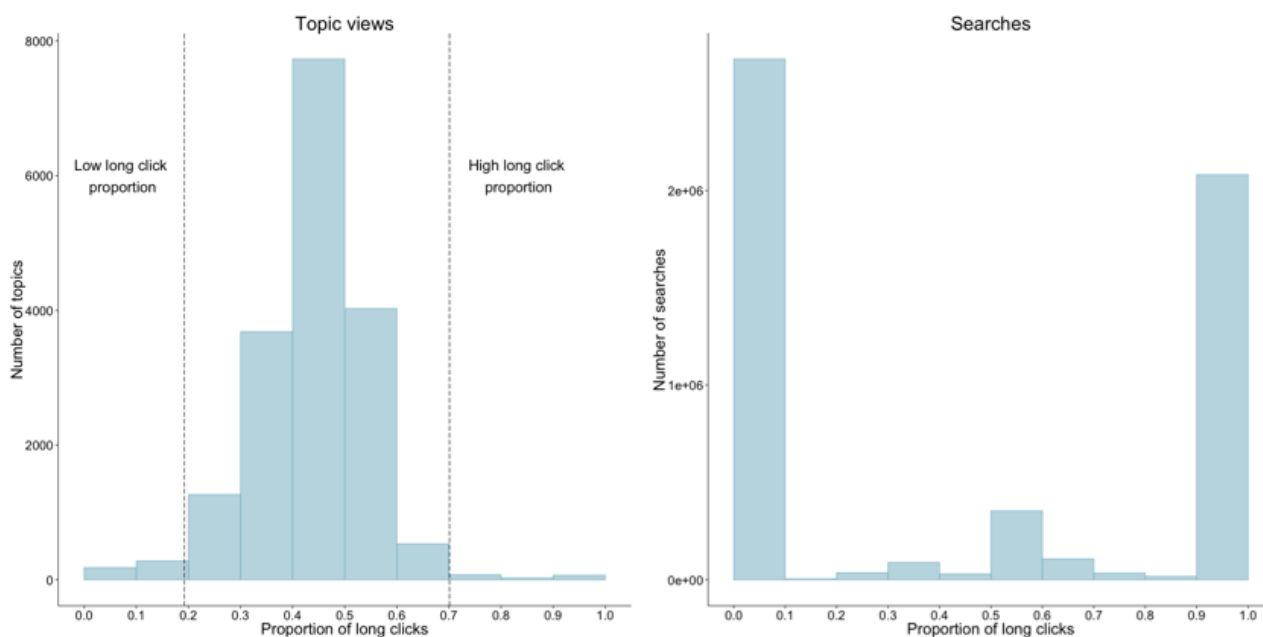
Table 1. Top 10 search terms (by volume) that did not initiate any long clicks for subsequent topic views.

Term	Frequency
gad7	187
adriamycin patient info	144
doxorubin patient info	109
phenochromocytoma	106
lab values in pregnancy	105
augment	86
probencid	79
dilaudid in pregnancy	71
parovirus	66
kidney ston	60

Table 2. Top 10 search terms (by volume) that initiated only long clicks for subsequent topic views.

Term	Frequency
abreva	3411
bullous myringitis	3392
subchorionic hemorrhage	1904
choroid plexus cyst	1270
cerefolin	1258
estropipate	1207
moexipril	1181
postpartum hypertension	1153
tinactin	1152
flucon	1114

Sequences of Searches and Topic Views

Table 3 lists the most frequently observed topic sequences of lengths 3 and 4, and the searches that initiated them. We found that topic sequences longer than 4 had a high proportion of topic "switching" (alternating repeatedly between two topics) and are thus not shown. These topic sequences had a common pattern: a topic view for a disease or condition concept was followed by topic views for drugs or treatments for the disease.

Table 3. The 10 most frequently observed topic view sequences of size 3 or 4, and the search terms that initiated them.

Topic view sequence (frequency)	Initiating search terms (frequency)
Cellulitis and erysipelas → Clindamycin: drug information → Clindamycin (systemic): drug information (14,463)	cellulitis (10,434), cellulitis treatment (2055), erysipelas (260), skin infection (219), facial cellulitis (171)
Onychomycosis → Terbinafine: drug information → Terbinafine (systemic): drug information (8234)	onychomycosis (3421), onychomycosis treatment (1433), onchomycosis (531), toenail fungus (524), nail fungus (476)
Onjunctivitis → Erythromycin: drug information → Erythromycin (ophthalmic): drug information (5938)	conjunctivitis (2738), conjunctivitis treatment (801), pink eye (603), bacterial conjunctivitis (422), conjunctivitis (292)
Treatment of clostridium difficile infection in adults → Metronidazole: drug information → Metronidazole (systemic): drug information (5923)	c diff (1947), c diff treatment (454), clostridium difficile treatment (439), clostridium difficile (416), c. diff (357)
Treatment of acute pancreatitis → Predicting the severity of acute pancreatitis → Calculator: Ranson criteria for pancreatitis prognosis → Calculator: Apache II scoring system (1219)	pancreatitis (731), acute pancreatitis (303), pancreatitis treatment (73), acute pancreatitis treatment (52), gallstone pancreatitis (23)
Management of acute exacerbations of chronic obstructive pulmonary disease → Management of infection in acute exacerbations of chronic obstructive pulmonary disease → Azithromycin: drug information → Azithromycin (systemic): drug information (565)	copd exacerbation (350), copd (98), copd exacerbation treatment (80), copd exacerbation antibiotics (24), copd exac (8)
Overview of diaper dermatitis in infants and children → Nystatin: drug information → Nystatin (topical): drug information → Nystatin (topical): pediatric drug information (522)	diaper rash (350), diaper dermatitis (71), diaper rash treatment (36), diaper candidiasis (19), candidal diaper rash (18)
Acute uncomplicated cystitis and pyelonephritis in women → Ciprofloxacin: drug information → Ciprofloxacin (ophthalmic): drug information → Ciprofloxacin (systemic): drug information (416)	uti (276), urinary tract infection (52), cystitis (31), uti treatment (26), pyelonephritis (16)
Overview of acute pulmonary embolism → Diagnosis of acute pulmonary embolism → Treatment of acute pulmonary embolism → Anticoagulation in acute pulmonary embolism (392)	pulmonary embolism (311), pe (50), pulmonary embolus (31)

Monitoring Response to a Food and Drug Administration Alert Using UpToDate and Online Media Activity

Finally, we used user search activity as seen in UpToDate logs to monitor the response of health care professionals to FDA alerts. Specifically, we measured the relative volume of searches about the antidepressant drug citalopram (trade name Celexa) prior to and following the August 24, 2011 FDA warning about the risk of abnormal heart rhythms when using citalopram. As described in the Methods, we compared the relative query volume in UpToDate to the relative volume of mentions for the drug in online media (see Figure 7 and Multimedia Appendix 1).

Online media showed a small surge in Celexa-related terms on the same day as the alert, and showed a sharp rise in the following days peaking at approximately 10 days. In contrast, the UpToDate query volume for Celexa did not increase until 10 days *after* the FDA alert but had sustained high query volumes for approximately the next 60 days, long after Celexa-related activity in the general online media returned to a baseline level.

The number of unique search terms that initiated Celexa-related topic views was much higher after the alert (see Table 4), but the proportion of long clicks for those topics was lower after the alert. This difference was even more pronounced when considering only the 2 months prior to and after the alert (data not shown). The average topic view duration for Celexa pages was also significantly lower after the FDA alert (842.31 seconds in the ~9 months before the alert compared to 744.36 seconds in the ~9 months after the alert).

Search specificity also increased following the FDA alert. The number of searches for "citalopram" or "Celexa" with terms "long qt," "heart," or "rhythm" was only 2 before the alert and 34 after the alert, demonstrating the effect of the FDA alert.

XSL•FO

RenderX

Figure 7. Scaled 7-day moving average of UpToDate query volume (red), 7-day moving average of media activity (blue), and raw media activity volume (grey) related to Celexa prior to and following the August 24, 2011 FDA warning (date indicated by the green dotted line).
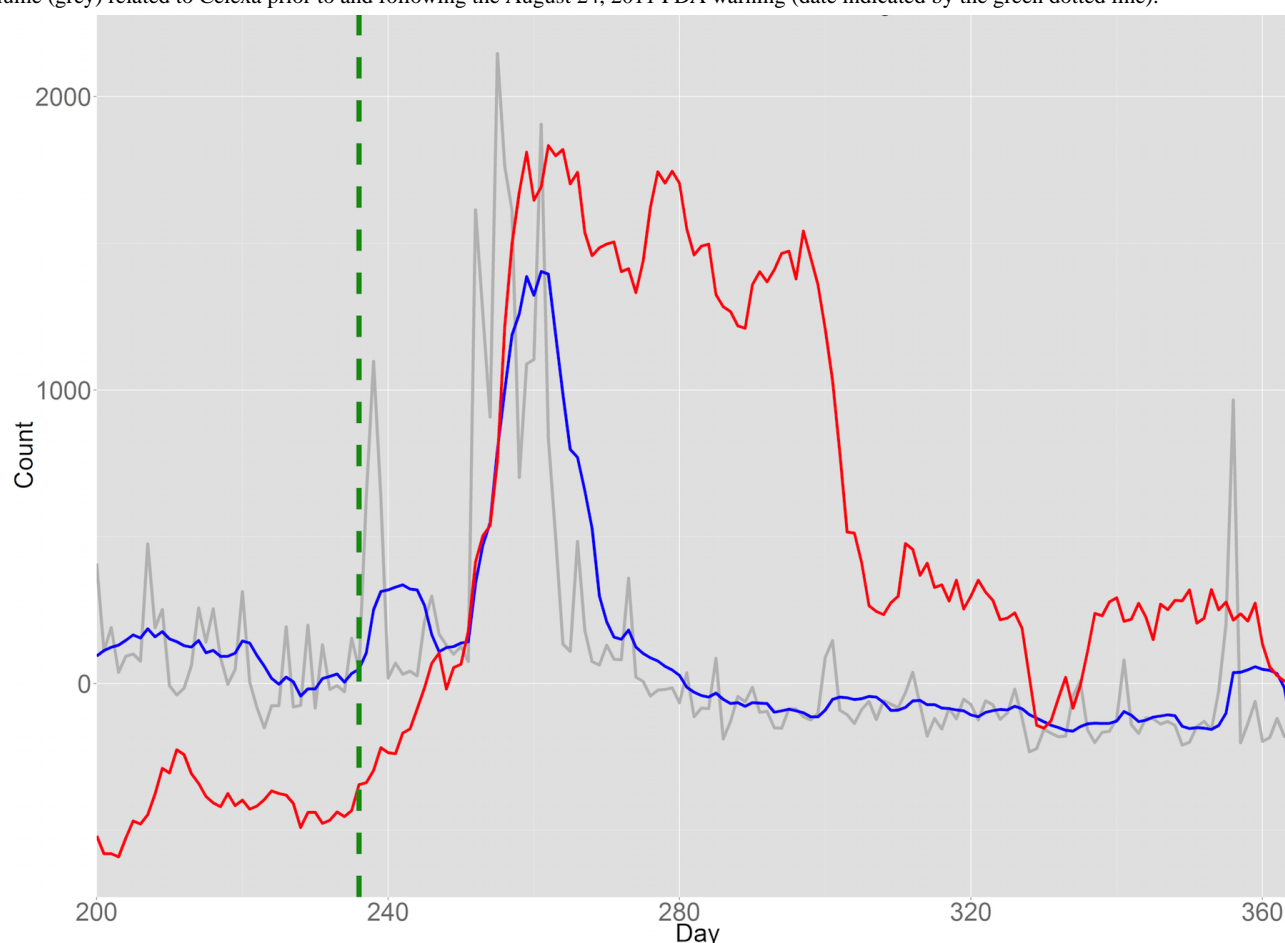
Table 4. Number of unique search terms that initiated Celexa (citalopram) topic views, and the proportion of long clicks for those topics before and after the August 24, 2011 FDA warning.

Topic	Number of unique search terms		Proportion of long clicks	
	Before alert	After alert	Before alert	After alert
Citalopram: Drug information	489	683	0.376	0.354
Citalopram: Patient drug information	183	303	0.262	0.247
Citalopram: Pediatric drug information	43	63	0.395	0.301

Discussion

Principal Findings

This is the first study profiling the online search behavior of medical professionals using a dedicated evidence-based medical information resource. Our findings can inform changes in the design of resources such as UpToDate. For example, search terms that only resulted in short clicks (indicating that the results were not useful to the searcher) potentially identify the need for new content or new search terms, while search terms and topics that have only high proportions of long clicks represent well-covered topics. Searches that have a seasonal variation in volume, such as searches for influenza and related medications, could be highlighted in the appropriate season to make the content more accessible. Such analyses can improve the utilization of the content by taking into account user

behavior—an activity that consumer Web companies routinely undertake.

Recent work on mining Web logs in the medical domain has been concerned with pharmacovigilance using consumer logs [13,16,32]. These studies use statistical methods to detect significant single- and multi-drug adverse event association signals from consumer search logs, taking into consideration the length of time between searches for drugs and adverse events and using known drug-induced adverse events as gold standards to evaluate the performance of their methods. For example, an analysis of consumer search logs showed that evidence of a multi-drug adverse event association between hyperglycemia, pravastatin, and paroxetine could be detected by quantifying the disproportionality of searches about hyperglycemia occurring with searches about both pravastatin and paroxetine, as compared to its co-occurrence with searches for only one of the drugs [32]. A second study demonstrated that combining data

from the FDA Adverse Event Reporting System (FAERS) with consumer search logs improved accuracy of adverse drug event detection by 19%, compared with using either source alone [16]. Our findings in this work and in a recent preliminary study [33] indicate that incorporating search logs of health care professionals for pharmacovigilance is a promising approach.

The uptake of the 2011 FDA alert for citalopram was clearly reflected in UpToDate logs, demonstrating the potential of this data source as a means to assess the efficacy and measure uptake of FDA alerts by health care professionals. This finding is also supported by previous work demonstrating the analysis of UpToDate usage logs to monitor influenza epidemics [24]. In combination with a recently published time-indexed reference set of adverse drug reactions extracted from recent FDA label changes and warnings [34], UpToDate logs could enable a large-scale analysis of physician response to FDA label changes and alerts. Our data extract (from 2011-2012) does not overlap with the time-indexed label changes. However, with access to recent data, such an analysis has the potential to inform the FDA of the effectiveness of their alerts more broadly.

Search log analysis may offer additional opportunities for surveillance by measuring changes in search volume about organ systems, diseases, and drugs over time, as well as by quantifying the relationship between changes in search volume and associated events. Such surveillance can monitor infectious disease outbreaks or watch for changes in the prevalence of health conditions that are a significant burden on the health care system. Development of methods to monitor such changes is possible and has been previously demonstrated by the use of consumer Internet search logs to predict health care utilization, detect flu outbreaks, and to track prescription drug use. The use of health professional search logs to improve upon such use cases is a research area with the potential to improve public health through earlier warning of disease outbreaks and to improve drug safety surveillance by monitoring physician response to FDA communications to assess their efficacy.

Limitations

There are several limitations to our approach. While the location, time, and associated user license of UpToDate searches and topic views are known, we do not have information on the identity of UpToDate users. We expect that the vast majority of licensed UpToDate users are health care professionals (ie, medical doctors, nurse practitioners, and/or researchers), but it is possible that some of the logs capture UpToDate use by patients or other types of consumers. Similarly, a unique session identifier relates searches and topic page views, but it is possible that within a single session there were multiple users with distinct information needs and behavior. Relying on the raw

logs, we are unable to identify user switching within sessions, and as a result may have associated searches and topic views that were actually performed by different users and are therefore unrelated. However, research has been dedicated to developing methods for automatically determining session boundaries in Web log data (including user switching)—for example, in work by Göker and He [35] and Murray et al [36]—which may be used to address this potential shortcoming. Also, as noted in the Methods, we ignored topic view events that were the final event in a session because it was not possible to calculate page view duration for those topic views. This reduced the number of topic view events available for analysis by 22%; the remaining 78% of topic views spanned 63% of all sessions. Finally, our findings are based on a relatively short surveillance period of 2 years—an analysis of logs from a longer period could reveal new associations or associations of differing strength.

More generally, data mining approaches applied to Web-scale search data may have methodological shortcomings. In early 2013 it was found that the Google Flu Trends system was overestimating flu prevalence, predicting values much higher than estimates from the Centers for Disease Control. Such inaccuracies may occur when methods are not recalibrated to adjust for temporal fluctuations that have an external cause, such as media coverage of the unusual 2012-2013 flu season resulting in more flu-related searches [37]. An analysis of Google Flu Trends data found that its prediction errors from week to week were correlated to each other and exhibited seasonality [38], suggesting that additional confounding variables (including changes to Google search algorithms themselves) may be partly responsible for changes in observed search term prevalence. Such potential confounders should be considered when interpreting our FDA alert findings, but could not be included in our analysis because we do not have access to UpToDate search engine features or ranking algorithms.

Conclusions

Our results demonstrate that mining UpToDate search logs offers unique insight into the information-seeking behavior of health care professionals, and the relationship between this behavior and health care utilization associated with disease states. Our results allow us to understand the information needs of health professionals in their day-to-day practice and the relationship between search terms and topic views—a large fraction of which include a disease or condition concept followed by drugs or treatments for the disease. Lastly, we were able to use UpToDate to quantify the uptake of the FDA alert for a serious drug adverse event, illustrating a novel use of analyzing search behavior to monitor responses to FDA alerts at a national level.

Conflicts of Interest

Howard Strasberg is an employee of Wolters Kluwer.

Multimedia Appendix 1

Scaled 7-day moving average of UpToDate query volume (red), 7-day moving average of media activity (blue), and raw media activity volume (grey) related to Celexa for all of 2011-2012, prior to and following the August 24, 2011 FDA warning (date indicated by the green dotted line).

[PNG File, 792KB - jmir_v17i8e204_app1.png]

References

1. Fox S. Pew Research Center. Washington, DC: Pew Internet & American Life Project; 2011 May 12. The social life of health information URL: http://www.pewinternet.org/2011/05/12/the-social-life-of-health-information-2011/ [accessed 2015-03-11] [WebCite Cache ID 6WxPUyoVr]

2. Fox S, Duggan M. Pew Research Center. Washington, DC: Pew Internet & American Life Project; 2013 Jan 15. Health online 2013 URL: http://www.pewinternet.org/2013/01/15/health-online-2013/ [accessed 2015-03-10] [WebCite Cache ID 6WxPbiHqx]

3. Masters K. For what purpose and reasons do doctors use the Internet: a systematic review. Int J Med Inform 2008 Jan;77(1):4-16. [doi: 10.1016/j.ijmedinf.2006.10.002] [Medline: 17137833]

4. Clarke MA, Belden JL, Koopman RJ, Steege LM, Moore JL, Canfield SM, et al. Information needs and information-seeking behaviour analysis of primary care physicians and nurses: a literature review. Health Info Libr J 2013 Sep;30(3):178-190. [doi: 10.1111/hir.12036] [Medline: 23981019]

5. Google/Manhattan Research. 2012 Jun. Screen to script: the doctor's digital path to treatment URL: https://www.thinkwithgoogle.com/research-studies/the-doctors-digital-path-to-treatment.html [accessed 2015-03-11] [WebCite Cache ID 6WxWKrYXP]

6. Kourouthanassis PE, Mikalef P, Ioannidou M, Pateli A. Exploring the online satisfaction gap of medical doctors: an expectation-confirmation investigation of information needs. Adv Exp Med Biol 2015;820:217-228. [doi: 10.1007/978-3-319-09012-2_15] [Medline: 25417027]

7. Ginsberg J, Mohebbi MH, Patel RS, Brammer L, Smolinski MS, Brilliant L. Detecting influenza epidemics using search engine query data. Nature 2009 Feb 19;457(7232):1012-1014. [doi: 10.1038/nature07634] [Medline: 19020500]

8. McIver DJ, Brownstein JS. Wikipedia usage estimates prevalence of influenza-like illness in the United States in near real-time. PLoS Comput Biol 2014 Apr;10(4):e1003581 [FREE Full text] [doi: 10.1371/journal.pcbi.1003581] [Medline: 24743682]

9. Ofran Y, Paltiel O, Pelleg D, Rowe JM, Yom-Tov E. Patterns of information-seeking for cancer on the internet: an analysis of real world data. PLoS One 2012;7(9):e45921 [FREE Full text] [doi: 10.1371/journal.pone.0045921] [Medline: 23029317]

10. Harsha AK, Schmitt JE, Stavropoulos SW. Know your market: use of online query tools to quantify trends in patient information-seeking behavior for varicose vein treatment. J Vasc Interv Radiol 2014 Jan;25(1):53-57. [doi: 10.1016/j.jvir.2013.09.015] [Medline: 24286941]

11. Althouse BM, Allem J, Childers MA, Dredze M, Ayers JW. Population health concerns during the United States' Great Recession. Am J Prev Med 2014 Feb;46(2):166-170. [doi: 10.1016/j.amepre.2013.10.008] [Medline: 24439350]

12. 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] [Medline: 23597817]

13. White RW, Horvitz E. From health search to healthcare: explorations of intention and utilization via query logs and user surveys. J Am Med Inform Assoc 2014;21(1):49-55 [FREE Full text] [doi: 10.1136/amiajnl-2012-001473] [Medline: 23666794]

14. Yom-Tov E, White RW, Horvitz E. Seeking insights about cycling mood disorders via anonymized search logs. J Med Internet Res 2014;16(2):e65 [FREE Full text] [doi: 10.2196/jmir.2664] [Medline: 24568936]

15. Schuster NM, Rogers MA, McMahon LF. Using search engine query data to track pharmaceutical utilization: a study of statins. Am J Manag Care 2010 Aug;16(8):e215-e219 [FREE Full text] [Medline: 20690788]

16. White RW, Harpaz R, Shah NH, DuMouchel W, Horvitz E. Toward enhanced pharmacovigilance using patient-generated data on the internet. Clin Pharmacol Ther 2014 Aug;96(2):239-246 [FREE Full text] [doi: 10.1038/clpt.2014.77] [Medline: 24713590]

17. Chisholm R, Finnell JT. Emergency department physician internet use during clinical encounters. AMIA Annu Symp Proc 2012;2012:1176-1183 [FREE Full text] [Medline: 23304394]

18. Kritz M, Gschwandtner M, Stefanov V, Hanbury A, Samwald M. Utilization and perceived problems of online medical resources and search tools among different groups of European physicians. J Med Internet Res 2013;15(6):e122 [FREE Full text] [doi: 10.2196/jmir.2436] [Medline: 23803299]

XSL•FO
RenderX

19.   Mickan S, Tilson JK, Atherton H, Roberts NW, Heneghan C. Evidence of effectiveness of health care professionals using handheld computers: a scoping review of systematic reviews. J Med Internet Res 2013;15(10):e212 [FREE Full text] [doi: 10.2196/jmir.2530] [Medline: 24165786]

20.   Samwald M, Kritz M, Gschwandtner M, Stefanov V, Hanbury A. Physicians searching the web for medical question answering: a European survey and local user studies. Stud Health Technol Inform 2013;192:1103. [Medline: 23920877]

21.   UpToDate. URL: http://www.uptodate.com/home [accessed 2015-08-05] [WebCite Cache ID 6aaPtW2AH]

22.   Bonis PA, Pickens GT, Rind DM, Foster DA. Association of a clinical knowledge support system with improved patient safety, reduced complications and shorter length of stay among Medicare beneficiaries in acute care hospitals in the United States. Int J Med Inform 2008 Nov;77(11):745-753. [doi: 10.1016/j.ijmedinf.2008.04.002] [Medline: 18565788]

23.   Isaac T, Zheng J, Jha A. Use of UpToDate and outcomes in US hospitals. J Hosp Med 2012 Feb;7(2):85-90. [doi: 10.1002/jhm.944] [Medline: 22095750]

24.   Santillana M, Nsoesie EO, Mekaru SR, Scales D, Brownstein JS. Using clinicians' search query data to monitor influenza epidemics. Clin Infect Dis 2014 Nov 15;59(10):1446-1450. [doi: 10.1093/cid/ciu647] [Medline: 25115873]

25.   Bhat S, Bevans M, Sengupta S. Measuring users' Web activity to evaluate and enhance advertising effectiveness. J Advert 2013 May 31;31(3):97-106. [doi: 10.1080/00913367.2002.10673679]

26.   Danaher PJ, Mullarkey GW, Essegaier S. Factors affecting Web site visit duration: a cross-domain analysis. J Mark Res 2006 May;43(2):182-194. [doi: 10.1509/jmkr.43.2.182]

27.   Lependu P, Iyer SV, Fairon C, Shah NH. Annotation analysis for testing drug safety signals using unstructured clinical notes. J Biomed Semantics 2012;3 Suppl 1:S5 [FREE Full text] [doi: 10.1186/2041-1480-3-S1-S5] [Medline: 22541596]

28.   Lependu P, Liu Y, Iyer S, Udell MR, Shah NH. Analyzing patterns of drug use in clinical notes for patient safety. AMIA Jt Summits Transl Sci Proc 2012;2012:63-70 [FREE Full text] [Medline: 22779054]

29.   Leskovec J, Backstrom L, Kleinberg J. Meme-tracking and the dynamics of the news cycle. In: Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '09). New York, NY: ACM; 2009 Presented at: 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '09); 2009; Paris, France p. 497-506.

30.   Suen C, Huang S, Eksombatchai C, Sosic R, Leskovec J. NIFTY: a system for large scale information flow tracking and clustering. In: Proceedings of the 22nd International Conference on World Wide Web (WWW '13). Geneva, Switzerland: International World Wide Web Conferences Steering Committee; 2013 Presented at: 22nd International Conference on World Wide Web (WWW '13); May 13-17, 2013; Rio de Janeiro, Brazil p. 1237-1248.

31.   Spinn3r. URL: http://spinn3r.com/ [accessed 2015-03-11] [WebCite Cache ID 6WxPiH8Nx]

32.   White RW, Tatonetti NP, Shah NH, Altman RB, Horvitz E. Web-scale pharmacovigilance: listening to signals from the crowd. J Am Med Inform Assoc 2013 May 1;20(3):404-408 [FREE Full text] [doi: 10.1136/amiajnl-2012-001482] [Medline: 23467469]

33.   Odgers DJ, Harpaz R, Callahan A, Stiglic G, Shah NH. Analyzing search behavior of healthcare professionals for drug safety surveillance. Pac Symp Biocomput 2015;20:306-317 [FREE Full text] [Medline: 25592591]

34.   Harpaz R, Odgers D, Gaskin G, DuMouchel W, Winnenburg R, Bodenreider O, et al. A time-indexed reference standard of adverse drug reactions. Sci Data 2014 Nov 11;1:140043 [FREE Full text] [doi: 10.1038/sdata.2014.43] [Medline: 25632348]

35.   Göker A, He D. Analysing Web search logs to determine session boundaries for user-oriented learning. In: Brusilovsky P, Stock O, Strapparava C, editors. Adaptive Hypermedia and Adaptive Web-Based Systems. Berlin, Germany: Springer Berlin Heidelberg; 2000:319-322.

36.   Murray GC, Lin J, Chowdhury A. Identification of user sessions with hierarchical agglomerative clustering. In: Proceedings of the American Society for Information Science and Technology Annual Meeting. 2007 Oct Presented at: American Society for Information Science and Technology Annual Meeting; October 19-24, 2007; Milwaukee, WI p. 1-9. [doi: 10.1002/meet.14504301312]

37.   Bulter D. Nature.: Nature Publishing Group; 2013 Feb 13. When Google got flu wrong URL: http://www.nature.com/news/when-google-got-flu-wrong-1.12413 [accessed 2015-03-11] [WebCite Cache ID 6WxPnZ7ph]

38.   Lazer D, Kennedy R, King G, Vespignani A. Big data. The parable of Google Flu: traps in big data analysis. Science 2014 Mar 14;343(6176):1203-1205. [doi: 10.1126/science.1248506] [Medline: 24626916]

Abbreviations

CCS: clinical classification system
FAERS: Food and Drug Administration Adverse Event Reporting System
FDA: Food and Drug Administration
HCUP: Healthcare Cost and Utilization Project
ICD-9: International Classification of Diseases, 9th Revision
IQR: interquartile range
NIGMS: National Institute of General Medical Sciences

XSL•FO
RenderX

NIH: National Institutes of Health
NLM: National Library of Medicine
UMLS: Unified Medical Language System

XSL•FO
RenderX

Original Paper

Assessing Electronic Cigarette-Related Tweets for Sentiment and Content Using Supervised Machine Learning

Heather Cole-Lewis[1], MA, MPH, PhD; Arun Varghese[1], MEng, MPA; Amy Sanders[1], MA; Mary Schwarz[1]; Jillian Pugatch[1], MPH; Erik Augustson[2], MPH, PhD

[1]ICF International, Rockville, MD, United States

[2]National Cancer Institute, Tobacco Control Research Branch, Bethesda, MD, United States

Corresponding Author:
Erik Augustson, MPH, PhD
National Cancer Institute
Tobacco Control Research Branch
9609 Medical Center Drive
MSC 9761
Bethesda, MD, 20892
United States
Phone: 1 240 276 6774
Fax: 1 240 276 7907
Email: augustse@mail.nih.gov

Abstract

Background: Electronic cigarettes (e-cigarettes) continue to be a growing topic among social media users, especially on Twitter. The ability to analyze conversations about e-cigarettes in real-time can provide important insight into trends in the public's knowledge, attitudes, and beliefs surrounding e-cigarettes, and subsequently guide public health interventions.

Objective: Our aim was to establish a supervised machine learning algorithm to build predictive classification models that assess Twitter data for a range of factors related to e-cigarettes.

Methods: Manual content analysis was conducted for 17,098 tweets. These tweets were coded for five categories: e-cigarette relevance, sentiment, user description, genre, and theme. Machine learning classification models were then built for each of these five categories, and word groupings (n-grams) were used to define the feature space for each classifier.

Results: Predictive performance scores for classification models indicated that the models correctly labeled the tweets with the appropriate variables between 68.40% and 99.34% of the time, and the percentage of maximum possible improvement over a random baseline that was achieved by the classification models ranged from 41.59% to 80.62%. Classifiers with the highest performance scores that also achieved the highest percentage of the maximum possible improvement over a random baseline were Policy/Government (performance: 0.94; % improvement: 80.62%), Relevance (performance: 0.94; % improvement: 75.26%), Ad or Promotion (performance: 0.89; % improvement: 72.69%), and Marketing (performance: 0.91; % improvement: 72.56%). The most appropriate word-grouping unit (n-gram) was 1 for the majority of classifiers. Performance continued to marginally increase with the size of the training dataset of manually annotated data, but eventually leveled off. Even at low dataset sizes of 4000 observations, performance characteristics were fairly sound.

Conclusions: Social media outlets like Twitter can uncover real-time snapshots of personal sentiment, knowledge, attitudes, and behavior that are not as accessible, at this scale, through any other offline platform. Using the vast data available through social media presents an opportunity for social science and public health methodologies to utilize computational methodologies to enhance and extend research and practice. This study was successful in automating a complex five-category manual content analysis of e-cigarette-related content on Twitter using machine learning techniques. The study details machine learning model specifications that provided the best accuracy for data related to e-cigarettes, as well as a replicable methodology to allow extension of these methods to additional topics.

XSL•FO

RenderX

Introduction

As evidenced by the announcement of "vape" as the Oxford dictionary word of the year in 2014, electronic cigarettes (e-cigarettes) are relevant and of interest to the general public [1]. Although the topic is pervasive to the general public, there is no definitive scientific evidence on the safety or effectiveness of e-cigarettes [2]. Furthermore, there is limited evidence on public knowledge, attitudes, and behaviors related to e-cigarettes [3-5]. This information is critical to guiding the development of public health communication, policies, and interventions regarding e-cigarettes.

With 74% of online adults using some form of social media [6], the digital landscape is continuing to evolve, and social media platforms such as Twitter have become platforms for public discourse at a local and global level on a variety of topics and events, including health and politics [7]. These discussions generate a massive amount of data that represent unfiltered public opinion and provide a unique opportunity for social science and public health research, especially for rapidly evolving topics, such as e-cigarettes.

Behavioral science and public health researchers traditionally turn to surveys, focus groups, and in-depth interviews to explore a particular topic. However, these techniques often require considerable resources, such as time and money. Additionally, these methodologies are subject to biases related to querying a person in a research setting that can affect the validity of findings (eg, social desirability bias). Furthermore, the landscape on some topics evolves rapidly and thus requires research that can be conducted quickly and with minimal resources in order to ensure that public health efforts are abreast of public knowledge, attitudes, and behaviors. Twitter analysis is currently a very active research area that offers an abundance of data for the behavioral and social sciences. The examination of this data can uncover trends in knowledge, attitudes, and behavior; inform public health and public policy; and pave the way for interventions delivered via social media, especially in the case of tobacco use and cessation [8-12]. The breadth of social media data available allows researchers to circumvent the aforementioned issues and explore opportunities to analyze this data, thus giving rise to infodemiology and infoveillance—the analysis of Internet content and electronic data sources to identify health-related trends and disease outbreaks [13]. This field may also be referred to as digital epidemiology or digital disease detection [14].

In 2009, Google Flu Trends aggregated search engine queries to track influenza activity in the United States, which strongly correlated with official surveillance data from the Influenza-like Illness Surveillance Network [15]. Google Flu Trends was able to detect regional outbreaks before conventional Centers for Disease Control and Prevention surveillance systems, bringing infoveillance and infodemiology closer to true real-time public health surveillance [16]. However, the methodology used was subject to biases and its influenza incidence predictions have often fallen short of reality [17]. Similarly, Chew and Eysenbach used infodemiology methods to monitor Twitter trends surrounding H1N1 in 2009, which included the adoption of the

WHO-recommended term "H1N1" over "swine flu", the use of retweets to disseminate information (versus opinions and personal experiences), and an early attempt at implementing automated analysis to monitor and analyze tweets in real-time [18]. Paul and Dredze [19] found that Twitter data correlate with public health metrics and knowledge such as syndromic surveillance to identify flu outbreaks, sentinel surveillance to identify the correlation of geographic behavioral risk factors and disease such as tobacco use and cancer, and the combination of both types of surveillance to identify seasonal allergies by geographic region, which lends further support for the use of social media data as cheaper and faster to obtain in comparison to survey data. Infodemiology via Twitter has also been used in studies tracking sentiment and informedness during natural disasters, misuse of antibiotics, and other public health issues and patterns [12,20-22].

Despite the widespread application of Twitter data for infodemiology, skeptics warn that the signal-to-noise ratios from sources like Twitter are very low and the demographics represented on Twitter represent younger voices with a larger proportion of minorities, making Twitter results less representative of the general public [23]. Nonetheless, as this area of research and application advances, there is increasing attention paid to the ethical challenges that arise from use of publicly available data and how to conduct research that acknowledges and addresses those challenges [14]. For example, Yin et al [24] developed a method to detect whether tweets originate from accounts run by individuals or accounts run by companies or organizations, thus providing opportunities to improve the signal-to-noise ratio when using Twitter data for infodemiology and infoveillance.

Given the growing opportunities for infodemiology in public health, it is important to continue improving upon existing methodologies for behavioral science and public health in order to increase accuracy and efficiency, as well as determining how best to utilize computational techniques to support traditional public health methods. Recent studies have implemented manual content analyses of data obtained via Twitter in order to assess public sentiment on the emerging topic of e-cigarettes [25] and the extent of e-cigarette marketing via Twitter [26]. Building on this, our research team developed a five-category coding scheme to classify and identify trends in public conversations about e-cigarettes based on information culled from Twitter over a 1-year period.

This research extended the science on previous manual content analyses for e-cigarettes because the five categories included in the manual content analysis were crafted specifically to inform public health communication, intervention, and research by focusing not only on sentiment about e-cigarettes and content of messages, but also important details such as characterization of the speaker. Findings from the manual content analysis and subsequent correlational analyses study revealed trends in e-cigarette conversations via Twitter. Results showed not only sentiment of tweets, but also the type of Twitter user discussing various categories of content, and how the conversation (and types of Twitter users driving that conversation) shifted over time. For instance, advertising and promotion-related tweets were the single largest content theme category, followed by

policy-related and then health-related tweets. Additionally, everyday users of Twitter generated a greater percentage of marketing-related tweets than retailers or tobacco companies, and everyday users of Twitter were also the top producers of tweets demonstrating first-person use or intent to use e-cigarettes with e-cigarette use.

While findings from this study are informative, the e-cigarette landscape continues to grow and thus, the manual content analysis would need to be replicated over time in order to continue to discover trends. Manual annotation of data requires considerable resources and time [27]. There is an opportunity to utilize computational methods such as machine learning to enhance and extend traditional public health methods. However, it is critical to explore the best computational models to support the task of automating the classification of Twitter data.

The purpose of this study was to determine feasibility of using computational natural language processing-based supervised machine learning techniques to replicate findings of a five-category manual content analysis of Twitter data related to e-cigarettes by using the manually coded data as a training set to train machine learning algorithms. This research builds on that of the tri-axial coding scheme used by Myslín et al [25]; however, the current study is unique in that it uses a more complex five-category coding scheme. Findings from the current study stand to provide insight into specific methodological considerations (eg, type of classification algorithm, size of word grouping analysis unit, and amount of information necessary) that enhance the performance of computational models designed to identify specifics of e-cigarette-related content on Twitter such as relevance, type of Twitter user, sentiment towards e-cigarettes, and more. Study findings stand to inform the development of public health-related infodemiology tools that may be deployed retrospectively and in real-time to explore public opinion on rapidly developing topics such as e-cigarettes.

Methods

Overview

In this study, supervised machine learning was used to build predictive classification models that assess Twitter data for a range of e-cigarette–related factors. Multiple classification models were created that varied by underlying machine learning classification technique and word-grouping units (ie, n-grams). Performance of classification models was assessed using 10-fold cross validation. Additionally, adequacy of sample size for manually coded content was determined by plotting model performance against varying sample sizes to build learning curves.

Data Collection and Manual Annotation

The corpus of tweets that formed the basis of this analysis was acquired from Gnip, a provider of historical Twitter data. Strategic keywords were used to collect historic tweets potentially related to e-cigarettes between May 1, 2013, and May 1, 2014. Keywords were selected by building on keyword lists used for similar research in the literature and adapted based on information of interest for the purposes of this study (Multimedia Appendix 1) [25]. Gnip provided all tweets meeting the keyword search during the time frame, which yielded 3.7 million tweets. Manual content analysis was conducted for a randomly selected subset of tweets, thus creating the dataset to be used in this analysis. Tweets were coded according to a codebook developed based on previous literature and adapted for the purposes of this study [12,18,25]. Six analysts independently coded a subset of 250 of the same tweets until an acceptable interrater reliability score was reached for each of the five categories detailed in Table 1. Interrater reliability was determined using the Fleiss' kappa statistic and a score of at least 0.64 was obtained for each category, indicating substantial or good agreement [28,29]. Definitions of the categories (relevance, sentiment, user description, genre, theme) can be found in Table 1 and Multimedia Appendix 2.

A total of 17,098 tweets were coded for relevance, of which 10,128 (59.23%) were found to be relevant and interpretable and therefore coded for the additional categories of sentiment, user description, genre, and theme. Of the 6970 non-relevant tweets, 2384 (34.20%) were found to be entirely non-relevant, whereas the remainder were retweets with no additional context, conversations without context, or duplicated tweets from a user account that had since been suspended or was primarily being used for spam or unwanted solicitations.

Table 1. Supervised machine learning-based e-cigarette tweet classification categories (interrater reliability score for manual annotation).

Classification (Fleiss' kappa)	Labels
	Relevant
	Subcategory: retweet with no additional information
	Subcategory: original tweets that were part of a conversation and require greater context to be interpreted
	Subcategory: duplicated tweets from a user account that had since been suspended or was primarily being used for spam or unwanted solicitations
Relevance[a]: Identifies tweets that are related to e-cigarettes (0.70)	Not relevant
Sentiment[b]: Indicates whether the stance in the tweet is positive, neutral, or negative towards e-cigarettes and users of e-cigarettes (0.65)	Positive
	Neutral
	Negative
User description[b]: Characterizes the sender of the tweet based on information gleaned from the user profile (0.66)	Celebrity
	Government
	Foundations or organizations
	Reputable news source
	Everyday people
	E-cigarette community movement
	Retailers
	Tobacco company
	Bots/hacked
Genre: Represents the format of the tweet (0.64)	Information
	First person e-cig use or intent
	Second/third person experience
	Personal opinion
	Marketing
	News/update
Theme: Refers to the topical domain of the content in the tweet (0.65)	Cessation
	Health and safety
	Underage usage
	Craving
	Other substances
	Illicit substance use in e-cigs
	Policy or government
	Parental use of e-cigs
	Advertisement/promotion
	Flavors

[a]Binary version of this category was created in addition to multiclass version for the purposes of the analysis.

[b]Categories were mutually exclusive and thus analyzed as multiclass.

Tweet Classification Model Construction

Machine learning classification models were built for each of the five categories (relevance, user description, sentiment, genre, theme). In order to determine the best performing classifier model, several variations of classification techniques and word-grouping units (n-gram) were used (see Multimedia Appendices 3 and 4). In order for the classification model to distinguish relevant tweets from non-relevant tweets, the entire dataset of manually coded 17,098 tweets was used. All relevant tweets (10,128) were used to build classification models for the sentiment, user description, genre, and theme.

A mathematical representation of the tweet corpus was created based on the term frequency inverse document frequency

transformation, which was preceded by the removal of stopwords and tokenization of text features established on count-based vectorization. For the final models, no attempt was made to reduce the feature space by using feature selection algorithms because exploratory analyses suggested no significant gain in performance and a potential decrease in predictive accuracy from implementing feature selection.

Three machine learning classification techniques—each based on alternative underlying statistical pattern recognition philosophies—were tested for each classifier: Naïve Bayes, k-Nearest Neighbors, and Support Vector Machines. In addition, word groupings (n-grams) ranging from unigrams to 5-grams were used to define the feature space for each classifier. The key attributes of the classification techniques used in this analysis are discussed in Multimedia Appendix 3 [30,31].

Assessment of Model Performance and Sample Size

The preferred measure of predictive performance of the classification models implemented and reported in this analysis was the accuracy score, defined as the percentage of observations that were correctly classified in the validation dataset. This method was chosen due to its simplicity in interpretation, measure of overall effectiveness of a classification model [32], neutrality with respect to the weighting of false positive and false negatives, ease of comparison with other studies in the field, and suitability for multiclass variables.

The analysis implemented 10-fold cross validation as a means to avoid bias in the estimation of the accuracy score. This involved dividing the manually classified data into 10 groups, iteratively using combinations of nine distinct groups to fit the model and the remaining group to validate the performance of the model, and averaging the predictive performance score.

Performance scores were evaluated for each of the three classification techniques described above for feature spaces described by n-grams between unigrams and 5-grams. A total of 15 classification models were thus evaluated for each classifier (3 classification techniques x 5 n-gram specifications=15 classification models). Classifier refers to the categorical labels that were assigned during the manual annotation process that the machine classification models seek to correctly label. Classifiers resulting from mutually exclusive categories (ie, user description, sentiment) were analyzed as

multiclass (could assume one of many class values) (see Table 1). Classifiers resulting from non-mutually exclusive categories—multilabel categories (eg, genre, theme) where one tweet could be assigned any one or more of many class values—were each assessed in terms of binary prediction performance for each constituent class (could assume only one of two class values). The relevance category was analyzed in the form of both binary and multiclass classifiers, given that data for subcategorization of the relevance category was also available. A total of 20 classifiers were assessed, each with 15 classification models.

Random accuracy baselines were computed for each binary and multiclass classifier to provide a point of performance comparison. The random baseline reflects how well a classification model would perform based on pure guesswork combined with knowledge of the true occurrence fraction of each class.

We assessed sample size adequacy by sequentially including 20%, 40%, 60%, 80%, and 100% of manually coded tweets and plotted a learning curve to visually examine where, if at all, the improvement in performance score begins to level off. As noted earlier, we also quantitatively assessed the feasibility of feature selection based on the chi-square method in improving efficiency and accuracy for a limited number of classifiers. This exploratory analysis concluded it was better to proceed without feature selection for the final models. The Python programming language version 2.7, in particular the Scikit Learn library version 0.15.1, was used for these analyses.

Results

The predictive performance scores from the supervised machine learning–based analyses are presented in Table 2. The table reports accuracy scores for the best performing classification model, with consideration to the best performing classification technique and best word-grouping unit (n-gram) for each classifier. As a method of normalization, accuracy scores were additionally evaluated in terms of the percentage improvement achieved of the maximum possible improvement over the random baseline. A more complete table of results, which includes all classification models (all combinations of classification techniques and n-grams evaluated for each classifier), is included in Multimedia Appendix 4.

Table 2. Supervised machine learning-based e-cigarette tweet classification performance results.

Classifier labels	Best n-gram	Accuracy score	% achieved of possible improvement over random baseline
Relevance category[a]	1	0.75	57.25
Relevance	1	0.94	75.26
User description[a]	2	0.68	41.59
Sentiment[a]	2	0.76	46.05
News	1	0.93	52.26
Info	4	0.86	41.75
Personal experience	2	0.84	50.17
Second person	2	0.92	47.09
Personal opinion	2	0.79	48.93
Marketing	1	0.91	72.56
Cessation	1	0.95	58.43
Health and safety	1	0.90	56.29
Underage usage	1	0.97	58.92
Craving	2	0.97	58.43
Other substances[b]	1	0.99	49.42
Illicit substances	2	0.98	48.24
Policy or government	1	0.94	80.62
Parental use	1	0.99	54.40
Ad or promotion	1	0.89	72.69
Flavor	1	0.97	62.52

[a]Classifiers were multiclass. All other categories were binary.

[b]k-nearest neighbors (kNN) was the best performing classification technique; for all other cases, linear support vector machine (SVM) was best.

Predictive performance scores for classification models ranged between 0.68 and 0.99 indicating that the models correctly labeled the tweets with the appropriate variables between 68.40% and 99.34% of the time, and the percentage of maximum possible improvement over a random baseline that was achieved by the classification models ranged from 41.59% to 80.62%. The average performance score was 0.90 and the average improvement over a random baseline was 56.64%. Classifiers with the highest performance scores that also achieved the highest percentage of the maximum possible improvement over a random baseline were Policy/Government (performance: 0.94; % improvement: 80.62%), Relevance (performance: 0.94; % improvement: 75.26%), Ad or Promotion (performance: 0.89; % improvement: 72.69%), and Marketing (performance: 0.91; % improvement: 72.56%).

All classifiers performed best using the linear support vector algorithm with the exception of Other Substances, which performed best with the k-Nearest Neighbors algorithm. The most appropriate word-grouping unit (n-gram) was 1 for the majority of classifiers. Twelve classification models performed best with a unigram sequence, while seven performed best with a bigram sequence, only one performed best with a four-gram sequence, and none performed best with tri-gram or 5-gram sequences.

Figures 1 and 2 display learning curves for a selection of classifiers. These curves indicate that performance continues to marginally increase with the size of the training dataset of manually annotated data but begins to level off at approximately 14,000 observations for relevance classification and approximately 8000 (relevant) observations for topic classification. However, even at low dataset sizes of 4000 observations, the performance characteristics are fairly sound. All classifiers, including those not displayed, followed this same pattern. Besides providing insight into how performance characteristics relate to sample size, these data suggest that an adequate training set of manually annotated data was deployed for the current analysis.

Figure 1. Learning curve for tweet relevance classification.
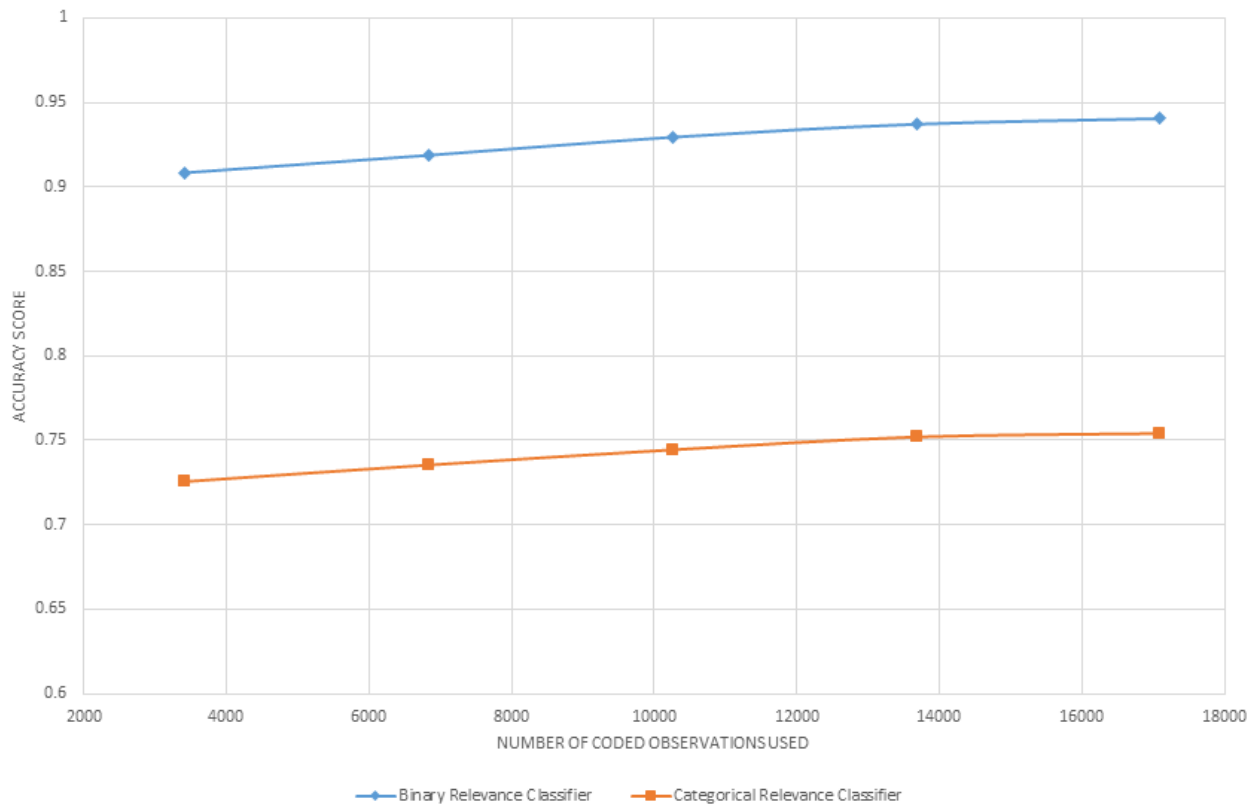
Figure 2. Learning curve for tweet topic classification.
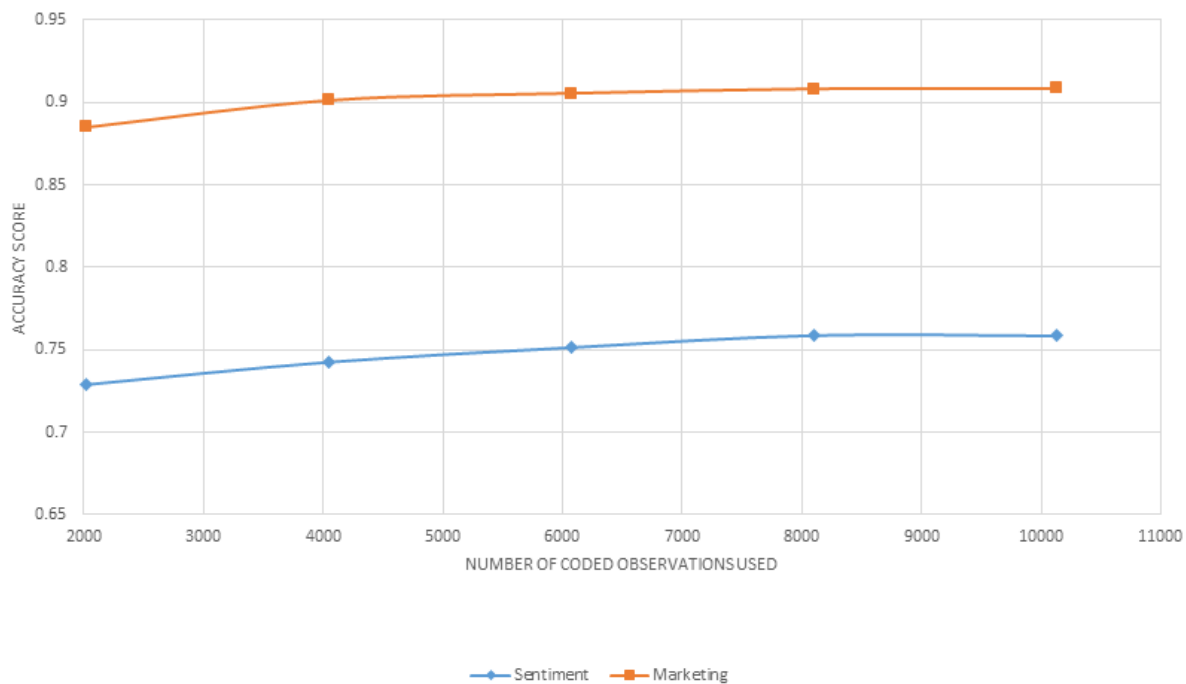
Discussion

Principal Findings

This work adds to the growing body of literature that highlights the importance of machine learning for large, language-based datasets of publicly available data. Using social media allows for exploration into conversations occurring outside of the traditional public health space, and machine learning provides an opportunity to keep abreast of these conversations in a more rapid fashion. The results of this study provide an example of the use of supervised machine learning methods to assess the vast social media landscape around e-cigarettes. This study used a five-category manually coded training set to train machine learning classification algorithms, thus categorizing e-cigarette-related content on Twitter with relative accuracy and detail. The findings provide insight into machine learning techniques that are most appropriate for assessing e-cigarette data around particular topics such as sentiment, speaker, and genre. The study provides a methodology that can be replicated to determine similar information about other public health-related trends and topics. Particular aspects of this methodology contribute to efforts to improve ethical use of Twitter data for public health, such as improving the signal-to-noise ratio [14]. For example, identification of the type of Twitter account allows information to be considered in context as opposed to considering all information disseminated via Twitter to be equal.

Of the classification techniques examined, linear support vector machines generally had the highest levels of predictive performance, which is consistent with the results of some previous text classification studies [25]. Unigrams were generally found to be the most successful word grouping for tweet classification, which is consistent with the short nature of tweets and their relative performance observed in previous studies [25,33].

The absolute performance scores from these models compare favorably to those reported in the literature for similar short text classification tasks [25,33]. For instance, Agarwal et al report classification accuracy ranging from 56.31% to 60.83% for a tertiary (positive, negative, neutral) sentiment analysis of manually annotated Twitter data [33], with a chance baseline of 33.33%, thus realizing from 34.47% to 41.25% of the total achievable improvement over the chance baseline. By comparison, our study realized 46.05% of the total achievable improvement over the chance baseline for a tertiary (positive, negative, neutral) sentiment classifier. For the 20 binary or multiclass variables considered in this study, this metric ranged between 41.59% and 80.61%.

Findings from learning curves assessing classification model performance by sample size of manually annotated data show that the sample size used in this study was sufficient to observe maximum performance of the classification models. Additionally, learning curve findings provide insight to future research to assess the optimal sample size of manually annotated data necessary to build such supervised machine learning algorithms. As Figueroa et al note, manual annotation (ie, manual content analysis) of data for supervised machine learning can be cumbersome; thus, knowledge such as that provided by this study can aid future researchers in making decisions related to optimization of the sample size for manually annotated training sets [27].

Further analyses of the rich dataset created as part of this work may contribute to the development of novel methods that could enhance the performance of automated surveillance tools. Unsupervised topic classification techniques could potentially be used in creative ways to improve the performance of the supervised learning classifier models. Machine learning–based image classification may add an additional dimension to automated surveillance tools assessing social media for insights and trends. The learning curve data developed for various combinations of algorithms and n-grams may be fitted with generalized mathematical functions that may potentially provide a basis for manual annotation sample size decision rules in other contexts.

Methods used in this study may find potential extensions in the development of automated social media infodemiology tools that could provide insight into the evolving social media landscape around e-cigarettes and other public health–related topics in real-time, thus providing valuable information for researchers, policy makers, and public health officials. Findings obtained from tools such as these could be used to inform interventions, policy, and communication strategies with up-to-date and time relevant information. Additionally, methods from this study can be used to support exploratory analysis and hypothesis generation on more nuanced aspects of a particular topic or to focus on a particular demographic or user group. Furthermore, the discoveries that this type of infodemiology yields could potentially be used to inform the public and test communication strategies to influence behavior in the interest of public health. Public health officials and researchers engaged in behavior change interventions, such as smoking cessation support, may even consider potentially developing custom applications based on the detection of and responses to particular tweet topics (eg, youth initiation of e-cigarette use).

Even though the Twitter analysis automation was successful, it is based on a manual content analysis, which may be subject to bias. Despite this, the coding scheme used for the manual content analysis was based on an existing scheme supported by the literature [25], and our coders had acceptable interrater reliability. For these reasons, we are confident in the validity of the results and the replicability of the methodology to further understand Twitter trends over time for e-cigarettes as well as other health topics. The implications of replication are far-reaching, especially as social media and other digital platforms continue to be a venue for unfiltered, real-time discourse.

Conclusion

There is great potential for using new forms of data in social science and public health. As the world transitions to sending and receiving information online, social media outlets like Twitter hold the potential to uncover real-time snapshots of personal sentiment, knowledge, attitudes, and behavior that is not as accessible, at this scale, through any other offline platform. This medium is arguably one of the quickest and

easiest means to identify trends or outbreaks and allow researchers, public health officials, and policy makers to respond in a collaborative way to inform the public about issues that can improve the quality and longevity of their lives. Despite the benefits of infodemiology, this nascent field presents unique ethical challenges as well as challenges that are inherent to the study of public health [21].

As we seek to understand the vast amount of data available via social media, social science and public health methodologies must adapt and use computational methodologies to enhance and extend research and practice. This study was successful in automating a complex five-category manual content analysis of e-cigarette–related Twitter content using machine learning techniques. The study detailed machine learning model specifications that provided the best accuracy for data related to e-cigarettes, as well as a replicable methodology to allow extension of these methods to additional topics. In the future, additional research will be needed to continue to enhance these methodologies and demonstrate their cost-effectiveness and feasibility as tools for intervention and real-time surveillance.

Acknowledgments

Conflicts of Interest

None declared.

Multimedia Appendix 1

Tweet filter keywords.

[PDF File (Adobe PDF File), 169KB - jmir_v17i8e208_app1.pdf]

Multimedia Appendix 2

Definitions of annotation categories.

[PDF File (Adobe PDF File), 191KB - jmir_v17i8e208_app2.pdf]

Multimedia Appendix 3

Description and key attributes of machine learning classification.

[PDF File (Adobe PDF File), 175KB - jmir_v17i8e208_app3.pdf]

Multimedia Appendix 4

Extended supervised machine learning-based e-cigarette tweet classification performances results.

[PDF File (Adobe PDF File), 191KB - jmir_v17i8e208_app4.pdf]

References

1. Oxford D. OxfordWords Blog. The Oxford Dictionaries Word of the Year is... vape URL: http://blog.oxforddictionaries.com/2014/11/oxford-dictionaries-word-year-vape/ [accessed 2015-02-02] [WebCite Cache ID 6W37Y0rxs]
2. Callahan-Lyon P. Electronic cigarettes: human health effects. Tob Control 2014 May;23 Suppl 2:ii36-ii40 [FREE Full text] [doi: 10.1136/tobaccocontrol-2013-051470] [Medline: 24732161]
3. Pearson JL, Richardson A, Niaura RS, Vallone DM, Abrams DB. e-Cigarette awareness, use, and harm perceptions in US adults. Am J Public Health 2012 Sep;102(9):1758-1766 [FREE Full text] [doi: 10.2105/AJPH.2011.300526] [Medline: 22813087]
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. [doi: 10.1016/j.addbeh.2014.10.019] [Medline: 25452058]
5. Wackowski OA, Delnevo CD. Smokers' attitudes and support for e-cigarette policies and regulation in the USA. Tob Control 2015 Jan 6:-. [doi: 10.1136/tobaccocontrol-2014-051953] [Medline: 25564281]
6. Pew Internet Project. Social networking fact sheet URL: http://www.pewinternet.org/fact-sheets/social-networking-fact-sheet/ [WebCite Cache ID 6W37lXQjB]
7. Paul MJ, Dredze M. A model for mining public health topics from Twitter. URL: http://www.cs.jhu.edu/~mpaul/files/2011.tech.twitter_health.pdf [accessed 2015-05-19] [WebCite Cache ID 6Ye8SzcYI]

8.   Emery SL, Vera L, Huang J, Szczypka G. Wanna know about vaping? Patterns of message exposure, seeking and sharing information about e-cigarettes across media platforms. Tob Control 2014 Jul;23 Suppl 3:iii17-iii25 [FREE Full text] [doi: 10.1136/tobaccocontrol-2014-051648] [Medline: 24935893]

9.   Pechmann C, Pan L, Delucchi K, Lakon CM, Prochaska JJ. Development of a Twitter-based intervention for smoking cessation that encourages high-quality social media interactions via automessages. J Med Internet Res 2015;17(2):e50 [FREE Full text] [doi: 10.2196/jmir.3772] [Medline: 25707037]

10. Rocheleau M, Sadasivam RS, Baquis K, Stahl H, Kinney RL, Pagoto SL, et al. An observational study of social and emotional support in smoking cessation Twitter accounts: content analysis of tweets. J Med Internet Res 2015;17(1):e18 [FREE Full text] [doi: 10.2196/jmir.3768] [Medline: 25589009]

11. Murphy J, Link M, Childs J, Tesfaye C, Dean E, Stern M. American Association for Public Opinion Research. Deerfield, IL: American Association for Public Opinion Research; 2014 May 28. Social Media in Public Opinion Research: Report of the AAPOR Task Force on Emerging Technologies in Public Opinion Research URL: http://www.aapor.org/AAPORKentico/AAPOR_Main/media/MainSiteFiles/AAPOR_Social_Media_Report_FNL.pdf [accessed 2015-08-17] [WebCite Cache ID 6aqxmRX7z]

12. Prochaska JJ, Pechmann C, Kim R, Leonhardt JM. Twitter=quitter? An analysis of Twitter quit smoking social networks. Tob Control 2012 Jul;21(4):447-449 [FREE Full text] [doi: 10.1136/tc.2010.042507] [Medline: 21730101]

13. Eysenbach G. Infodemiology and infoveillance: framework for an emerging set of public health informatics methods to analyze search, communication and publication behavior on the Internet. J Med Internet Res 2009;11(1):e11 [FREE Full text] [doi: 10.2196/jmir.1157] [Medline: 19329408]

14. Vayena E, Salathé M, Madoff LC, Brownstein JS. Ethical challenges of big data in public health. PLoS Comput Biol 2015 Feb;11(2):e1003904 [FREE Full text] [doi: 10.1371/journal.pcbi.1003904] [Medline: 25664461]

15. Cook S, Conrad C, Fowlkes AL, Mohebbi MH. Assessing Google flu trends performance in the United States during the 2009 influenza virus A (H1N1) pandemic. PLoS One 2011;6(8):e23610 [FREE Full text] [doi: 10.1371/journal.pone.0023610] [Medline: 21886802]

16. Carneiro HA, Mylonakis E. Google trends: a web-based tool for real-time surveillance of disease outbreaks. Clin Infect Dis 2009 Nov 15;49(10):1557-1564 [FREE Full text] [doi: 10.1086/630200] [Medline: 19845471]

17. Lazer D, Kennedy R, King G, Vespignani A. Big data. The parable of Google Flu: traps in big data analysis. Science 2014 Mar 14;343(6176):1203-1205. [doi: 10.1126/science.1248506] [Medline: 24626916]

18. Chew C, Eysenbach G. Pandemics in the age of Twitter: content analysis of Tweets during the 2009 H1N1 outbreak. PLoS One 2010;5(11):e14118 [FREE Full text] [doi: 10.1371/journal.pone.0014118] [Medline: 21124761]

19. Paul MJ, Dredze M. You are what you tweet: Analyzing Twitter for public health. 2011 Jul 17 Presented at: Fifth International AAAI Conference on Weblogs and Social Media; July 2011; Barcelona, Spain p. 17-21 URL: http://www.aaai.org/ocs/index.php/ICWSM/ICWSM11/paper/view/2880/3264%20

20. Lowe JB, Barnes M, Teo C, Sutherns S. Investigating the use of social media to help women from going back to smoking post-partum. Aust N Z J Public Health 2012 Feb;36(1):30-32. [doi: 10.1111/j.1753-6405.2012.00826.x] [Medline: 22313703]

21. Chou WS, Prestin A, Kunath S. Obesity in social media: a mixed methods analysis. Transl Behav Med 2014 Sep;4(3):314-323 [FREE Full text] [doi: 10.1007/s13142-014-0256-1] [Medline: 25264470]

22. Ayers JW, Althouse BM, Dredze M. Could behavioral medicine lead the web data revolution? JAMA 2014 Apr 9;311(14):1399-1400. [doi: 10.1001/jama.2014.1505] [Medline: 24577162]

23. Butler D. When Google got flu wrong. Nature 2013 Feb 14;494(7436):155-156. [doi: 10.1038/494155a] [Medline: 23407515]

24. 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; 2014 May 08 Presented at: 18th Pacific-Asia Conference on Knowledge Discovery and Data Mining; May 2014; Tainan, Taiwan p. 163-175.

25. Myslín M, Zhu S, Chapman W, Conway M. Using twitter to examine smoking behavior and perceptions of emerging tobacco products. J Med Internet Res 2013;15(8):e174 [FREE Full text] [doi: 10.2196/jmir.2534] [Medline: 23989137]

26. Huang J, Kornfield R, Szczypka G, Emery SL. A cross-sectional examination of marketing of electronic cigarettes on Twitter. Tob Control 2014 Jul;23 Suppl 3:iii26-iii30 [FREE Full text] [doi: 10.1136/tobaccocontrol-2014-051551] [Medline: 24935894]

27. Figueroa RL, Zeng-Treitler Q, Kandula S, Ngo LH. Predicting sample size required for classification performance. BMC Med Inform Decis Mak 2012;12:8 [FREE Full text] [doi: 10.1186/1472-6947-12-8] [Medline: 22336388]

28. Fleiss J, Levin B, Paik M. Statistical Methods for Rates and Proportions. New York, NY: Wiley & Sons; 2003.

29. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics 1977 Mar;33(1):159-174. [Medline: 843571]

30. Dasarathy B. NN concepts and techniques: An introductory survey. In: Nearest Neighbour Norms: NN Pattern Classification Techniques. Los Alamitos, CA: IEEE Computer Society Press; Dec 1990:1-30.

31. Webb A. Statistical Pattern Recognition, 2nd ed. Chichester, England: John Wiley & Sons; 2002.

32. Sokolova M, Lapalme G. A systematic analysis of performance measures for classification tasks. Information Processing & Management 2009 Jul;45(4):427-437. [doi: 10.1016/j.ipm.2009.03.002]

XSL•FO
RenderX

33. Agarwal A, Xie B, Vovsha I, Rambow O, Passonneau R. Sentiment analysis of Twitter data. In: Proceedings of the Workshop on Languages in Social Media. 2011 Jun 23 Presented at: Workshop on Languages in Social Media; June 23, 2011; Portland, Oregon p. 30-38.

XSL•FO
RenderX

<u>Original Paper</u>

Association of Online Health Information–Seeking Behavior and Self-Care Activities Among Type 2 Diabetic Patients in Saudi Arabia

Amr Jamal[1,2,3*], MD, SBFM, ABFM, MRCGP, MBI; Samina A Khan[1*]; Ahmed AlHumud[4]; Abdulaziz Al-Duhyyim[4]; Mohammed Alrashed[4]; Faisal Bin Shabr[4]; Alwalid Alteraif[4]; Abdullah Almuziri[4]; Mowafa Househ[5], MEng, PhD; Riaz Qureshi[2,3]

[1]College of Medicine, Medical Informatics and E-Learning Unit, Medical Education Department, King Saud University, Riyadh, Saudi Arabia

[2]College of Medicine, Family and Community Medicine Department, King Saud University, Riyadh, Saudi Arabia

[3]KSU Chair for Medical Education Research and Development, King Saud University, Riyadh, Saudi Arabia

[4]College of Medicine, King Saud University, Riyadh, Saudi Arabia

[5]College of Public Health and Health Informatics, King Saud Bin Abdulaziz University for Health Sciences, Ministry of National Guard Health Affairs (MNGHA), Riyadh, Saudi Arabia

[*]these authors contributed equally

Corresponding Author:
Amr Jamal, MD, SBFM, ABFM, MRCGP, MBI
College of Medicine
Family and Community Medicine Department
King Saud University
POBox 90714
Riyadh, 11623
Saudi Arabia
Phone: 966 114690822
Fax: 966 114690821
Email: amrjamal@ksu.edu.sa

Abstract

Background: Health information obtained from the Internet has an impact on patient health care outcomes. There is a growing concern over the quality of online health information sources used by diabetic patients because little is known about their health information–seeking behavior and the impact this behavior has on their diabetes-related self-care, in particular in the Middle East setting.

Objective: The aim of this study was to determine the online health-related information–seeking behavior among adult type 2 diabetic patients in the Middle East and the impact of their online health-related information–seeking behavior on their self-care activities.

Methods: A cross-sectional survey was conducted on 344 patients with type 2 diabetes attending inpatient and outpatient primary health care clinics at 2 teaching hospitals in Riyadh, Saudi Arabia. The main outcome measures included the ability of patients to access the Internet, their ability to use the Internet to search for health-related information, and their responses to Internet searches in relation to their self-care activities. Further analysis of differences based on age, gender, sociodemographic, and diabetes-related self-care activities among online health-related information seekers and nononline health-related information seekers was conducted.

Results: Among the 344 patients, 74.1% (255/344) were male with a mean age of 53.5 (SD 13.8) years. Only 39.0% (134/344) were Internet users; 71.6% (96/134) of them used the Internet for seeking health-related information. Most participants reported that their primary source of health-related information was their physician (216/344, 62.8%) followed by television (155/344, 45.1%), family (113/344, 32.8%), newspapers (100/344, 29.1%), and the Internet (96/344, 27.9%). Primary topics participants searched for were therapeutic diet for diabetes (55/96, 57%) and symptoms of diabetes (52/96, 54%) followed by diabetes treatment (50/96, 52%). Long history of diabetes, familial history of the disease, unemployment, and not seeking diabetes education were the most common barriers for online health-related information–seeking behavior. Younger age, female, marital status, higher

education, higher income, and longer duration of Internet usage were associated with more online health-related information–seeking behaviors. Most (89/96, 93%) online health-related information seekers reported positive change in their behaviors after seeking online health information. Overall odds ratio (OR 1.56, 95% CI 0.63-3.28) for all self-care responses demonstrated that there was no statistically significant difference between those seeking health-related information online and non–health-related information seekers. However, health-related information seekers were better in testing their blood glucose regularly, taking proper action for hyperglycemia, and adopting nonpharmacological management.

Conclusions: Physicians and television are still the primary sources of health-related information for adult diabetic patients in Saudi Arabia whether they seek health-related information online or not. This study demonstrates that participants seeking online health-related information are more conscious about their diabetes self-care compared to non–health-related information seekers in some aspects more than the others.

Introduction

The expansion of the Internet has enabled people all over the world to gain access to a substantial amount of information on a variety of topics related to health sciences, human sciences, literature, and history [1,2]. Today, in the era of information technology, diabetic patients have become more dependent on online sources to access health information ubiquitously, especially with the propagation of smartphones, tablets, and laptops. Health-related information on the Internet for diabetes encompasses thousands of websites, chat rooms, and support groups that can be accessed by health consumers [2,3]. The medical community has studied the positive effects that online health-related information can have on patients, especially diabetic patients [4-7]. Previously published studies related to health information-seeking behaviors of diabetic patients have also addressed the potential benefits of online health-related information accessibility for diabetic patients as they search for information and advice about symptoms, disorders, and their appropriate treatments for diabetes [4-7]. An increasing number of patients are searching online for health information related to diabetes. Many of these patients have low health literacy levels and may retrieve inaccurate, incomplete, or out-of-date health information. [2,8,9]. Despite potential risks associated with online health-related information, millions of people use the Internet to search for diabetes-related health information. A recent study on the health-related information-seeking behaviors of a diabetes online community found that users engaged in peer support, advocacy, self-expression, humor, sharing, and seeking diabetes information [10]. The study also reports on the potential risks for diabetic patients searching for health-related information, which includes misinformation and privacy risks. The study recommends that although the Internet provides opportunities for communication between diabetic patients and health care providers, more research is needed to investigate the impact of health-related information on diabetes self-care [10].

In Saudi Arabia, Internet usage is rapidly growing and already slightly more than half of the population is using Internet [11,12]. The practice of using the Internet to seek health-related information is also common among patients in Saudi Arabia [2]. One of the most common and disabling diseases that patients need health-related information on is type 2 diabetes mellitus. The prevalence of type 2 diabetes mellitus in Saudi Arabia is worrying because already 20% of the adult population has this disease and it is expected to exceed to 25% by 2035 [13,14]. A national multistage survey study conducted in 2013 on 10,735 Saudi participants aged 15 years or older reported a high prevalence of diabetes (13.4%). A large proportion (43.6%) of diabetic individuals were undiagnosed before and only 29.1% of those receiving treatment had controlled diabetes. In addition, 15.2% were borderline diabetic. These numbers are alarming because they indicate a total of 1,745,532 diabetic and 979,953 borderline diabetic Saudis [15]. According to the International Diabetes Federation (IDF), Saudi Arabia has the fastest rate of growth of diabetes among the Middle East and North Africa (MENA) countries and the seventh highest in the world [13]. However, despite the high penetration of the Internet in Saudi society, there is a scarcity of existing research on the effect of diabetes health-related information-seeking behavior and its impact on self-care. The purpose of this study is to (1) determine online health-related information-seeking behavior among Saudi adult patients diagnosed with type 2 diabetes and (2) evaluate the impact of online health-related information-seeking behavior among diabetic patients on their self-care.

Methods

Study Design

The data for the current study were derived from a hospital-based cross-sectional survey conducted on a convenience sample of adult Saudi male and female patients diagnosed with type 2 diabetes in an outpatient and inpatient setting.

Setting

The study was conducted at King Saud University Medical City, Riyadh, Saudi Arabia, from February 28 to the end of March 2013. King Saud University Medical City consists of 2 teaching hospitals, which are tertiary referral hospitals with major primary health care outpatient/inpatient departments and serves patients of all sociodemographic levels in Riyadh and other parts of

country. Thus, a person in Saudi Arabia can use their services freely without restriction to a specific catchment area.

Participants

The target population was patients aged 16 years or older diagnosed with type 2 diabetes according to the American Diabetes Association's *Standards of Medical Care in Diabetes* guidelines [16]. All medical staff caring for diabetic patients (eg, physicians, nurses, technicians) was excluded from the study. Additionally, those who were not fluent in Arabic or English language were also excluded.

Sample Size

The diabetic patients were enrolled from outpatient and inpatient areas of the previously mentioned hospitals. The single proportion formula was used to calculate the sample size with 95% confidence level and 5% confidence interval [2]. The total sample size targeted in data collection of the present study was 344 diabetic patients.

Data Collection

The study instrument was a structured questionnaire developed in both English and the Arabic language and was adapted from previous work. The questionnaire included the following sections: (1) demographic information, (2) general Internet usage, (3) online health-related information–seeking behavior, and (4) questions related to self-care. The survey questionnaire was pilot-tested on 20 hospital volunteers at King Khalid University Hospital, Riyadh, to determine participants' level of comprehension. The results of the pilot study have not been included in this paper. Our trained team of researchers completed the survey by interviewing the participants individually in their preferred language of Arabic or English. The survey was conducted between February 28 and March 31, 2013. The questionnaire was validated through the pilot study feedback. A Cronbach alpha of greater than .6 was also determined for the instrument reliability.

Data Analysis

The study data were collected and entered into a computer using standardized entry codes. For all tests, statistical significance was set at $P<.05$. Descriptive statistics were used to present means, standard deviations, and percentages. In addition, Student *t* test, *z* proportional test, and chi-square tests were employed to compare group variables between sexes, age groups, and other demographic variables. The relationships between demographic and Internet search for health-related information / self-care were assessed using binary unconditional multiple logistic regression analysis. The questionnaires were converted into binary data to run binary logistic regression. Adjusted odds ratios and the corresponding 95% confidence intervals (CIs) were calculated for each independent variable. Modeling was performed with the goal of selecting the most parsimonious and reasonable explanatory model that explained the relationship between independent and dependent variables. For bivariate analyses, all available data points were utilized. However, for multivariable analyses (logistic regression), a dataset was constructed that only had complete values for all relevant variables across the observations, thereby discarding the observations that had missing values for any of the variables involved in the regression analysis. This strategy was adopted to maintain comparability between models so that they could be developed from the same denominator. Factor subgroups were recombined for use in logistic regression analysis to prevent quasi-separation of cells resulting from compact cell sizes, which allowed the models to converge and yet provided for meaningful analyses. All analyses were conducted in SPSS version 21 (SPSS Inc, Chicago, IL, USA).

Ethics Statement

All participants were informed about the aim of the study and their verbal consent for participation was recorded. The study was approved by the Institutional Review Board at the College Of Medicine, King Saud University, Riyadh, Saudi Arabia, and was conducted in accordance with the declaration of Helsinki for Human Studies [3].

Results

Response Rate and Demographic Information

Of the 394 patients we approached to participate in the study, 344 (87.3%) completed the survey, whereas 37 of 394 (9.4%) participants decided not to take part in survey due to a concurrent scheduled physician appointment and 13 of 394 (3.3%) withdrew from the study due to lack of time to complete the survey.

Of the 344 diabetes patients who were interviewed, 255 (74.1%) were males. The overall mean age of participants was 53.5 (SD 13.8) years (males: mean 54.2, SD 14.1, range 16-84 years; females: mean 51.5, SD 12.9, range 19-80 years). Further demographic features are presented in Table 1.

Table 1. Sociodemographic details of samples based on age group and gender.

Sociodemographic characteristics	Total, n (%) N=344	Age band, n (%)			
		≥45 years n=255		<45 years n=89	
		Male n=192	Female n=63	Male n=63	Female n=26
Patient place					
Outpatient	302 (87.2)	169 (88.1)	55 (87)	54 (86)	24 (92)
Inpatient	42 (12.2)	23 (11.9)	8 (13)	9 (14)	2 (8)
Diabetes duration					
1-5 years	85 (24.7)	35 (18.2)	17 (27)	22 (35)	11 (42)
6-10 years	71 (20.6)	39 (20.3)	15 (24)	19 (30)	10 (39)
11-15 years	105 (30.5)	41 (21.4)	15 (24)	13 (21)	2 (8)
≥16 years	83 (24.1)	77 (40.1)	16 (25)	9 (14)	3 (12)
Are any of your first-degree relatives diabetic?					
Yes	267 (77.6)	143 (74.5)	52 (83)	50 (79)	22 (85)
No	77 (22.4)	49 (25.5)	11 (18)	13 (21)	4 (15)
Marital status					
Single	26 (7.6)	2 (1.0)	1 (2)	17 (27)	6 (23)
Married	291 (84.6)	187 (97.4)	42 (67)	46 (73)	16 (62)
Divorced	1 (0.3)	1 (0.6)	2 (3)	0 (0)	3 (12)
Widowed	0	2 (1.0)	18 (29)	0 (0)	1 (4)
Educational level					
Intermediate school or lower	126 (36.6)	67 (34.9)	50 (79)	4 (6)	5 (19)
High school	112 (32.5)	67 (34.9)	11 (17)	23 (37)	11 (42)
University	94 (27.3)	49 (25.5)	2 (3)	33 (52)	10 (39)
Postgraduate	12 (3.4)	9 (4.7)	0 (0)	3 (5)	0 (0)
Occupation					
Employed	125 (36.3)	60 (31.3)	9 (14.3)	46 (73)	10 (39)
Private business	40 (11.6)	27 (14.1)	5 (8)	8 (13)	0 (0)
Student	14 (4.1)	0 (0)	0 (0)	10 (16)	4 (15)
Unemployed	61 (17.7)	14 (7.3)	35 (56)	1 (2)	11 (42)
Retired	129 (37.5)	108 (56.3)	17 (27)	2 (3)	2 (8)
Household average monthly income (SR)					
<5000	13 (3.8)	43 (22.4)	39 (62)	17 (27)	12 (46)
5000-10,000	96 (27.9)	52 (27.1)	16 (25)	22 (35)	6 (23)
10,000-15,000	67 (19.5)	45 (23.4)	5 (8)	12 (19)	5 (19)
15,000-20,000	39 (11.3)	31 (16.2)	2 (3)	5 (8)	2 (8)
>20,000	30 (8.7)	21 (10.9)	1 (2)	7 (11)	1 (4)
Did you receive diabetes education?					
Yes	138 (40.1)	80 (41.7)	12 (19)	37 (59)	9 (35)
No	206 (59.9)	112 (58.3)	51 (81)	26 (41)	17 (65)
Do you use the Internet?					

Sociodemographic characteristics	Total, n (%) N=344	Age band, n (%)			
		≥45 years n=255		<45 years n=89	
		Male n=192	Female n=63	Male n=63	Female n=26
Yes	134 (38.9)	60 (31.2)	8 (13)	51 (81)	15 (58)
No	210 (61.0)	132 (68.8)	55 (87)	12 (19)	11 (42)

Only 134 of 344 (39.0%) of the interviewed patients were Internet users in general. The majority of Internet users were from younger age groups (Figure 1). Among those who used the Internet, 89.6% (120/134) had access to the Internet at home, 44.8% (60/134) had access at work, and 63.4% (85/120) had access on their mobile phone. Younger patients were more likely to be Internet users than older patients and Internet use declined with increasing age (Figure 1). Similarly, Internet use for health-related information was higher among younger participants. All females (23/23, 100%) who were already using the Internet in general were also using it to seek health-related information, whereas only 65.8% (73/111) of male participants who were already using Internet in general were also using it for health-related information (phi –0.286391, P<.001) (Figure 2). But this pattern was lower for the additional age band and no female aged 60 years or older used the Internet for health-related information. Most Internet users (105/134, 78.4%) reported effective (good/very good) skills of Internet searching

(Table 2). The majority of online health-related information seekers (76/96, 79%) reported that their main source of information was still their physician. However, only 96 of 134 Internet users (71.6%) reported using the Internet for health-related information with a mean search frequency of 6.4 (SD 9.9) times per month and a median of 2 (IQR 1-5) times per month (Table 2). Among all surveyed participants (N=344), most of the non-Internet users reported that their primary source of health-related information was their physician (216/344, 62.8%) followed by television (155/344, 45.1%), family (113/344, 32.8%), and newspapers (100/344, 29.1%), whereas 66 of 344 (19.2%) stated none for any health-related information sources (Figure 3). It was observed that online health-related information seekers used a variety of health-related information sources and these sources were preferred significantly by online health-related information seekers compared to the seekers of non–health-related information (Figure 4).

Table 2. Characteristics of Internet users included in the study.

Characteristics	Total, n (%) n=134	Age band, n (%)	
		≥45 years n=68	<45 years n=66
I have access to the Internet			
At home	120 (89.6)	59 (87)	61 (92)
At work	60 (44.8)	32 (47)	28 (42)
On my mobile phone	85 (63.4)	42 (62)	43 (65)
Frequency of use per month			
0-5 times	75 (55.9)	42 (62)	33 (50)
6-10 times	8 (6.0)	5 (7)	3 (5)
≥11 times	13 (9.7)	5 (7)	8 (12)
Rating of searching skills			
Very good	51 (38.1)	22 (32)	29 (44)
Good	54 (40.3)	25 (37)	29 (44)
Fair	22 (16.4)	15 (22)	7 (11)
Poor	7 (5.2)	6 (9)	1 (2)
Have you used the Internet to search for health information?			
Yes	96 (71.6)	52 (77)	44 (67)
Never	38 (28.4)	16 (24)	22 (33)

XSL•FO
RenderX

Figure 1. Distribution frequency of Internet use among diabetic patients by age.
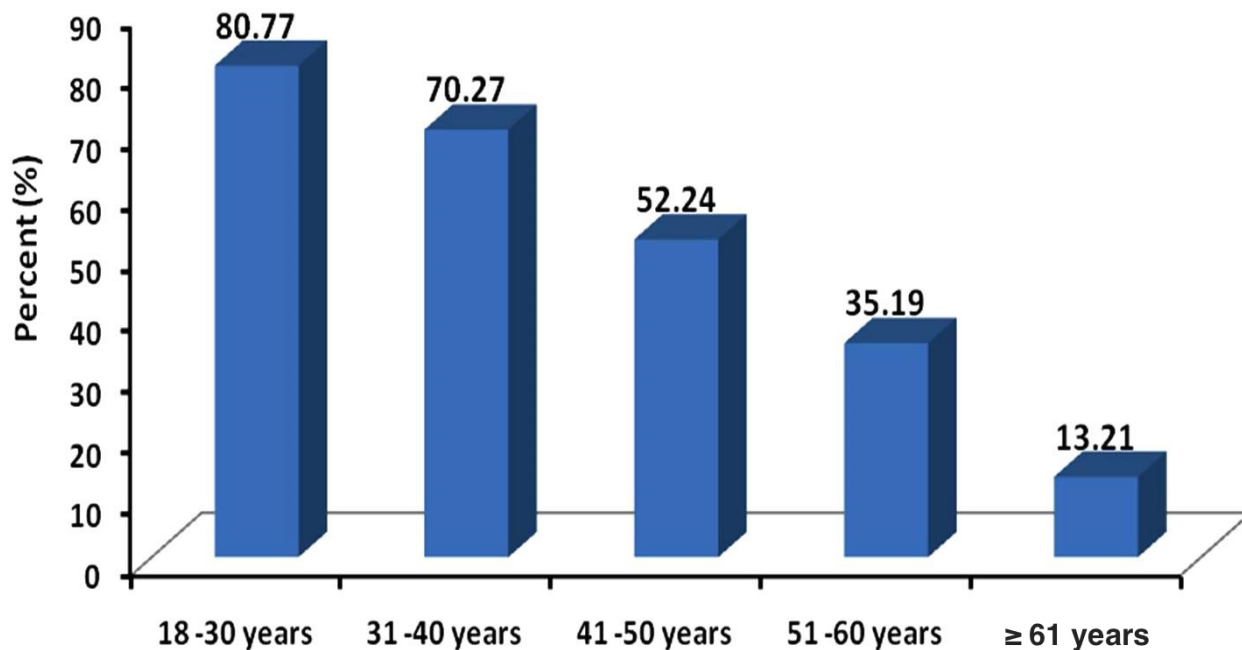
Figure 2. Distribution frequency of health information seekers among diabetic patients by age and gender.
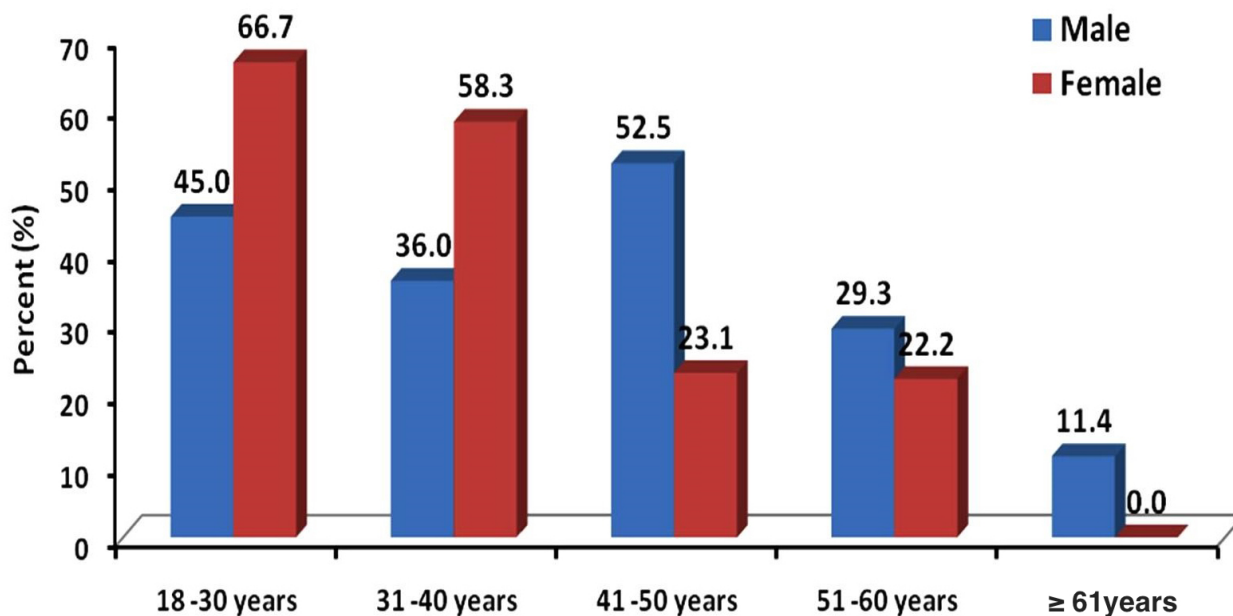
Figure 3. Sources of health-related information for online and nononline health-related information seekers.
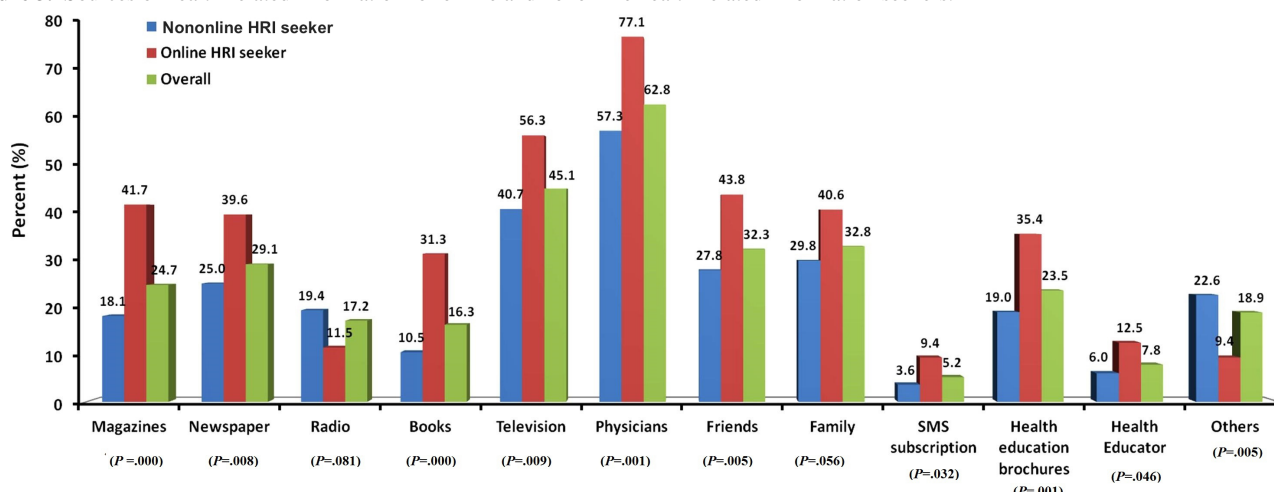
Figure 4. Types and frequency of Internet sites visited by diabetes patients to search for health-related information.
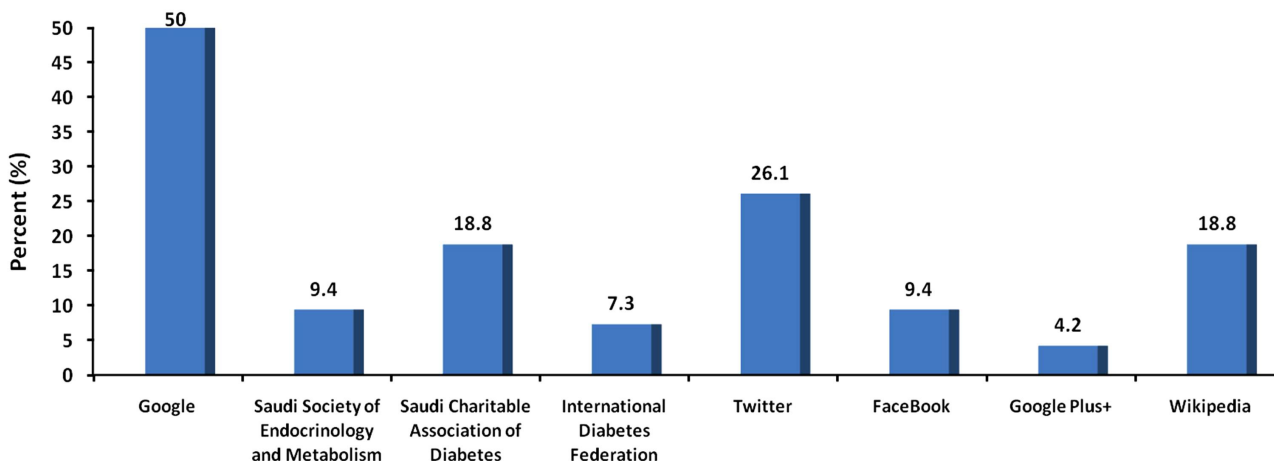
General Internet Usage

The majority of online health-related information seekers used Google as the primary search engine to look for the health information (94/96, 98%) followed by Bing (2/96, 2%), whereas no one selected the Yahoo search engine. The criterion for how they selected the link from the search result list was related to the perceived compatibility with the words they searched (62/96, 65%), followed by first link in the search result (25/96, 26%), and lastly URL type (eg, org/gov/edu/com) (12/96, 13%). The most frequent website used by study participants was Google (48/96, 50%) followed by Twitter (25/96, 26%); the Saudi

Charitable Association of Diabetes and Wikipedia shared the same percentage (18/96, 19%) (Figure 4). The majority of online health-related information seekers were searching for general health knowledge (64/96, 67%) followed by treatment of health problems (47%, 45/96) (Table 2). Whereas for diabetes-related information, the primary topics the participants were looking for were related to therapeutic diet for diabetes and symptoms of diabetes, followed by diabetes treatment and causes of diabetes (Table 3). Most of the online health-related information seekers (89/96, 93%) reported positive change in their behaviors after seeking online health information.

Table 3. Characteristics of online health-related information seekers included in the study.

Characteristics of health-related information seeker	Total, n (%) n=96	Age band, n (%)	
		≥45 years n=52	<45 years n=44
Where do you start looking for information on the Internet?			
MSN	2 (2)	1 (2)	1 (2)
Google	94 (98)	51 (75)	43 (65)
Yahoo	0 (0)	0 (0)	0 (0)
Bing	0 (0)	0 (0)	0 (0)
What was the primary reason you asked/looked for health information during the past year?			
Manage health	45 (47)	20 (29)	25 (38)
Diagnose health problem	29 (30)	15 (22)	14 (21)
Info about diseases prevention	26 (27)	17 (25)	9 (14)
For general health knowledge	64 (67)	36 (53)	28 (42)
For health and wellness info	33 (34)	25 (37)	8 (12)
Identify symptoms of health condition	11 (11)	7 (10)	4 (6)
For answering a specific question	21 (3)	10 (15)	11(17)
Other (please specify)	3 (3)	2 (3)	1 (2)
What was the primary type/topic of health information that you asked about or looked for during the past year?			
What is diabetes	38 (40)	19 (28)	19 (29)
Symptoms of diabetes	52 (54)	27 (40)	25 (38)
Causes of diabetes	48 (50)	23 (34)	25 (38)
Diagnosis of diabetes	24 (25)	13 (19)	11 (17)
Side effects of diabetes medications	34 (35)	21 (31)	13 (20)
Lifestyle management	31 (32)	16 (24)	15 (23)
Diabetes treatment	50 (52)	30 (44)	20 (30)
Specific health condition info	9 (9)	9 (13.2)	0 (0)
Therapeutic diet for diabetes	55 (57)	29 (43)	26 (39)
Therapeutic diet to lose weight	35 (36)	20 (29)	15 (23)
Complications of diabetes	31 (32)	19 (28)	12 (18)
Other (please specify)	2 (2)	2 (3)	0 (0)
After seeking health information and finding this information, did your health behavior change for the better?			
Yes	89 (93)	46 (88)	43 (98)
No	7 (7)	6 (12)	1 (2)

Online Health-Related Information Seeking Behavior

Logistic regression was performed to assess the impact of a number of sociodemographic factors on online health-related information seeking behavior (Table 4). The model contained 10 independent variables (sex, age band, marital status, education, income, occupation, diabetes duration, diabetes education, genetic run of diabetes). The full model containing all predictors was statistically significant ($P<.001$) indicating that the model was able to distinguish between respondents who used Internet for health-related information and correctly classified 72.3% cases. The strongest predictor was found to be

age band; those using the Internet for health-related information were more than 2.59 times (OR 2.59, 95% CI 0.88-3.15) more likely to be among the lower age group participants. Similarly, marital status and education level were also associated factors for seeking health-related information. Duration of diabetes and familial history of diabetes were negative predictors, suggesting that patients with longer duration of diabetes and a family history of diabetes were less likely to use the Internet for health-related information.

The odds ratio of 0.458 (95% CI 0.119-1.761) for occupation was less than 1, indicating that those who were either retired or

unemployed were 55% less likely to use the Internet for health-related information. Even those who reported to have exposure to diabetes education were 4.3% less likely to use the Internet for health-related information compared to nonexposed patients. The mean duration of Internet usage for health-related information seekers and non–health-related information seekers was 7.45 (SD 4.2) times per month and no statistical difference was found comparing health-related information seekers giving mean duration of 7.62 (SD 4.3) times per month and non–health-related information seekers (mean 7.62, SD 4.3 times per month) using Student *t* test on basis of Internet usage. Overall age, gender, marital status, education, income, and diabetes education were found to be important factors associated with online health-related information behavior.

Table 4. Logistic regression (N=344 full case data only) modeling odds for nononline health-related information seekers versus online health-related information seekers with sociodemographic details.

Sociodemographic characteristics	Total, n (%) N=344	Not health-related information seekers, n (%) n=248	Health-related information seekers, n (%) n=96	OR (95% CI)	*P*
Gender					
Female	89 (25.9)	66 (26.6)	23 (24)	1.459 (0.737, 2.888)	.29
Male	255 (74.1)	182 (73.4)	73 (76)	Ref	—
Age band					
≤45 years	89 (25.9)	45 (18.1)	44 (46)	2.593 (0.918, 7.323)	.001
>45 years	257 (74.7)	203 (81.9)	54 (56)	Ref	—
Marital status					
Never married	26 (7.6)	15 (6.0)	11 (12)	2.036 (0.477, 8.686)	.34
Married	318 (92.4)	233 (94.0)	85 (89)	Ref	—
Education					
Less than high school	238 (69.2)	194 (78.2)	44 (46)	0.403 (0.219, 0.740)	.52
University degree	98 (28.5)	54 (21.8)	44 (46)	Ref	—
Household monthly income (SR)					
≤10,000	216 (62.8)	164 (66.1)	52 (54)	0.458 (0.119, 1.761)	.33
>10,000	127 (36.9)	84 (33.9)	43 (45)	Ref	—
Occupation					
Unemployed	179 (52.0)	127 (51.2)	52 (54)	1.081 (0.325, 3.595)	.90
Employed	181 (56.6)	121 (48.8)	60 (63)	Ref	—
Diabetes education					
No	202 (58.7)	166 (66.9)	36 (38)	0.406 (0.232, 0.709)	.002
Yes	122 (35.5)	82 (33.1)	40 (42)	Ref	—
Duration of diabetes					
≤10 years	168 (48.8)	112 (45.1)	56 (58)	0.957 (0.900, 117)	.16
>10 years	176 (51.2)	136 (54.8)	40 (42)	Ref	—
Family history of diabetes					
No	267 (77.6)	183 (73.8)	84 (88)	0.438 (0.126, 1.528)	.20
Yes	77 (22.4)	65 (26.2)	12 (13)	Ref	—

Impact of Health-Related Information Users and Nonusers on Self- Care

Another logistic regression model was performed to assess the impact of seeking online health-related information on self-care among diabetic patients. Table 5 presents the logistic regression analysis or odds of health-related information seekers and nonseekers of self-care health information. The overall model was significantly better in explaining the relationship between online health-related information seekers and self- care. Overall, 4 self-care–related activities were significant factors in the model. Although most of the factors by themselves were not significant factors, they were retained in the model because of their contribution to the overall model as demonstrated by the likelihood ratio test. Removing these factors from the model changed the smaller model significantly from the one that included these factors; therefore, they were retained in the model (Table 5). Out of 12 self-related activities questions, 7 activities

showed higher positive association with online health-related information seekers. The strongest association of online health-related information seekers were observed for "their blood glucose check by themselves" and it was found that this check was 4.63 times (OR 4.63, 95% CI 1.86-11.56) more likely to be performed by online health-related information seekers compared to the health-related information nonseekers. With regards to testing for glucose, 28.6% (71/248) of non–health-related information seekers could test it themselves, whereas 93% (89/96) of health-related information seekers could test it themselves (*P*=.001).

For high blood glucose, 68% (65/96) of online health-related information seekers knew what to do correctly, whereas only 44.4% (110/248) of non–health-related information seekers did

(*P*=.003). Additionally, online health-related information seekers (33/96, 34%) were more likely to be aware of the importance of exercise and diet on the management of diabetes than non–health-related information seekers (42/248, 16.9%, *P*=.006) (Table 5). There was no statistically significant difference between online health-related information seekers and nonseekers for ophthalmologist and family physician checkups and performing diabetic foot self-exams.

The overall odds ratio (OR 1.56, 95% CI 0.63-3.28) of all self-care questionnaire responses demonstrated that there was no statistically significant difference between those seeking health-related information online and traditional health-related information seekers.

Table 5. Logistic regression to predict impact of health-related information seekers and nonseekers on self-health care activities.

Self-care health characteristics	Total, n (%) N=344	Not health-related information seeker, n (%) n=248	Health-related information seeker, n (%) n=96	OR (95% CI)	P
Do you test your blood glucose (sugar) by yourself?					
Yes	77 (22.4)	71 (28.6)	89 (92.7)	4.63 (1.86, 11.56)	.001
No	264 (76.7)	177 (71.4)	7 (7.3)	Ref	—
How do you treat high blood glucose?					
Inject insulin/avoid eating/others	305 (88.7)	219 (88.3)	87 (90.6)	1.21 (0.43, 2.75)	.87
Don't know	36 (10.5)	28 (11.3)	9 (9.4)	Ref	—
Do you wear a medical identification bracelet or necklace?					
Yes	11 (3.2)	9 (3.6)	2 (2.1)	1.51 (0.20, 5.60)	.95
No	330 (95.9)	239 (96.4)	94 (97.9)	Ref	—
Do you have a glucagon kit at home for severe lows (blood glucose)?					
Yes	67 (19.5)	41 (16.5)	27 (28.1)	0.11 (0.63, 1.13)	.12
No	274 (79.7)	207 (83.5)	69 (71.9)	Ref	—
In case of severe high blood glucose (sugar)					
Do something	175 (50.9)	110 (44.4)	65 (67.7)	0.14 (0.24, 1.32)	.003
No/don't know	166 (48.3)	136 (54.8)	31 (32.3)	Ref	—
What type of treatment that you take to manage diabetes?					
Insulin injections					
Yes	157 (45.6)	123 (49.6)	36 (37.5)	2.46 (1.29, 4.69)	.006
No	184 (53.5)	125 (50.4)	60 (62.5)	Ref	—
Diabetes pills					
Yes	250 (72.7)	179 (72.2)	72 (75.0)	1.33 (0.67, 2.66)	.41
No	91 (26.5)	69 (27.8)	24 (25.0)	Ref	—
Nonpharmacological treatment (exercises and diet)					
Yes	75 (21.8)	42 (16.9)	33 (34.4)	0.94 (0.41, 1.29)	.006
No	266 (77.3)	206 (83.1)	63 (65.6)	Ref	—
Do you test for ketones in the urine?					
Yes	169 (49.1)	119 (48.0)	52 (54.2)	0.95 (0.56, 1.62)	.85
No	172 (50.0)	129 (52.0)	44 (45.8)	Ref	—
Approximately how often do you visit a doctor for your diabetes?					
1-4 times a year	328 (95.3)	237 (95.6)	94 (97.9)	1.27 (091, 2.40)	.36
Never	13 (3.8)	11 (4.4)	2 (2.1)	Ref	—
How many times do you usually go to ophthalmologist for checking your eyes' retina?					
1-3 visit a year	206 (59.9)	143 (57.7)	65 (67.7)	1.27 (0.38, 1.21)	.19
Never	135 (39.2)	105 (42.3)	31 (32.3)	Ref	—
How many times do you usually check your feet by yourself?					
Daily/once a week	260 (75.6)	188 (75.8)	75 (78.1)	1.59 (0.80, 3.13)	.18
Never	81 (23.5)	60 (24.2)	21 (21.9)	Ref	—
Overall impact					
Yes (more health conscious)	173 (50.4)	123 (49.8)	58 (60.5)	1.56 (0.63, 3.28)	.33
No active health response	168 (48.7)	124 (50.1)	38 (39.5)	Ref	—

Discussion

Principal Findings

This study evaluated the extent of Internet use when searching for health-related information among type 2 diabetes patients visiting inpatient and outpatient clinics at 2 large public University Hospitals in Riyadh, Saudi Arabia. The main finding of this study shows that among diabetic patients, the primary sources of health-related information were physicians followed by television, friends, and magazines. Approximately one-quarter of the sampled diabetes patients were using the Internet for health-related information. The major factors associated with online health-related information seeking behavior were age, gender, marital status, educational level, and exposure to diabetic health education. Overall, these study findings have demonstrated that those seeking online health-related information were more conscious about their diabetes self-care compared to non–health-related information seekers.

In Saudi Arabia, Internet usage has increased rapidly over the past 10 years from less than 3% in 2000 to 60% in 2014; today, the trend shows signs of leveling off, particularly among younger age groups [11,12]. Among the surveyed diabetes patients in this study, only 39.0% (134/344) reported having Internet access and 27.9% (96/344) were online health-related information seekers. The Internet usage among diabetic patients was slightly lower compared to the overall national usage data. Additionally, the percentage of Internet use for health-related information in this study is lower than similar studies performed previously in Saudi Arabia, United States, Switzerland, Italy, and India, although these studies were not performed on patients with specific diseases [2,17-20]. Perhaps because English is not the primary language in Saudi Arabia, knowledge of the English language could be a factor influencing how diabetes patients search for health-related information on the Internet. The current findings also show that majority of the participants search only in Arabic. A greater number of participants were searching in Arabic, which is the native language in Saudi Arabia, whereas only 45% (43/96) of the health-related information seeker participants searched in both Arabic and English. Generally, most of the diabetic patients were elderly [13]. This has been revealed by a high mean age (mean 53.47, SD 13.8 years) in this study's participants. Generally, relatively older participants are not frequent users of the Internet and other digital devices and, even if they do, they may face some obstacles due to lack of searching skills [21-24].

These study results suggest that the physicians, followed by television, family, newspapers, and the Internet are the primary sources of health-related information. Despite the increasing consumer autonomy with the advent of the Internet, the physician remains one of the most preferred sources of health information in the new media environment, suggesting that more doctors need to explore the Internet as a viable medium for communicating with their patients. These results are concurrent with previous reports [1,2,25].

In a US study conducted in 2007, only 19% of online health-related information seekers searched information at least once a week, whereas one-third of the participants used the Internet at least once a month and individuals with a higher education level searched the Internet for health-related information more than any others [17]. However, in a study from Switzerland, the majority of participants searched for health-related information less than once per month [19]. According to some other studies, age, knowledge of the English language, acuteness or chronicity of the disease, and its severity were associated with the frequency of medical searches on the Internet [18,19]. The use of the Internet to search for health-related information decreased in males with age, whereas the highest rate of Internet use for health-related information among females was younger than age 45 years. Males with higher education and single individuals search for health-related information more than others [18]. Similarly, in this study, Internet searches for health-related information were associated with age, marital status, gender, education, diabetes education, and income. Unmarried individuals, females, those with higher education, and a higher income were found to be the more frequent online searchers for health-related information. These results show how education is an important factor with regards to the use of the Internet in searching for health-related information. This was expected because educated individuals and those who can afford digital devices or computers have greater access to the Internet. The current study suggests that a major factor associated with lower health-related information use among patients with diabetes is their age. Results of the multivariable model used in this study show that this is a continuous effect (OR 2.593, 95% CI 0.918-7.323), with Internet usage decreasing with increasing age. Similar observations were reported in previous studies [2,26]. Thus, improving older adults' access to and comfort with Internet technologies is central to implementing technology-based solutions to help patients manage their diabetes better.

Online solutions, such as Web portals are increasingly touted as a strategy to improve communication, provide support, and connect to needed services and information for patients with diabetes [21,26,27]. Because the prevalence of diabetes increases with age, particularly type 2 diabetes, the target population with the most to gain from health information technology is older patients. The diabetic patients in this study had a variety of motives and used different websites for their searches. However, the majority of the participants stated that they start their search for health information in Google because they do not know where else to go and the Google generator gives them options to find the links. In addition to Google searches, some patients were also using social sites, such as Twitter, Saudi Charitable Association of Diabetes, Wikipedia, and Facebook, to seek online health information. The use of social media technologies for online health-related information also allows the online social media users to create, distribute, share information, and to consult online rankings and reviews of independent websites [25]. Most of the participants in this study reported that the primary reasons for searching health information were general health knowledge, management of health, and for health and wellness info, respectively. For diabetes-related issues, they were searching for information on diet and symptoms of diabetes.

The current study found that patients searching for health information (89/96, 93%) report positive effects on their self-care behavior to managing their diabetes. Similar findings were reported previously for chronic illness and diabetes patients [15,21,22]. A greater number of online health-related information-seeking participants reported that they were more likely to test their blood glucose (sugar) by themselves as compared to the non-health-related information seekers. Likewise, online health-related information seekers who were testing their blood glucose were more aware about the methods for treating low blood glucose compared to the non-health-related information seekers. Only a fraction of participants who were non-health-related information seekers were aware of how to manage high blood glucose by doing something (eg, take pills, insulin injection, drink water) to alleviate their symptoms. On the other hand, the majority of online health-related information seekers were able to manage themselves for dealing with high blood glucose by doing exercise and following a strict diet, and additionally taking pills, insulin injections, and drinking more water during high blood glucose episodes. The majority of online health-related information seekers were also more likely to be aware of how to manage their disease themselves and to visit an ophthalmologist and a family physician regularly for checkups. The health-related information seekers also checked their feet by themselves on a regular basis. Overall, this study's findings have demonstrated that those seeking health-related information are more aware and conscious about their health self-care and were able to manage most of the diabetes-related self-care themselves compared to the non-health-related information seekers. Thus, this study's findings are concurrent with previous reports that explaining the correct uses of the Internet to obtain health-related information can lead to better patient awareness for treatment decisions and increased patient satisfaction, resulting in better medical outcomes and improved self-care behavior [17,27]. Additionally, online health-related information improves the physician-patient relationship and increased patient satisfaction [2]. Similarly, most online health-related information seekers in this study significantly reported that physicians are their primary source of health-related information compared to the nonusers. They usually discuss their disease with their physician to know more about self-care. Interestingly, the majority of the participants were not aware about the quality of the websites and the basic information they provide related to their query. Perhaps much of the information on the Internet could be misleading because anyone can claim medical expertise. Most of the time, information on the Internet is incomplete, out of date, and the public might not be able to select valid information [2,19]. To manage and provide better information to patients, health policy makers should prepare guidelines and strictly keep their eyes on fraudulent and misleading diabetes-related information. To empower patient's knowledge and self-care, physicians should ask their patients if they are using the Internet to obtain diabetes-related health-related information and provide them with reliable/trusted website information.

Implications for Practice

This study found that 39.0% (134/344) of participants use the Internet. As a result of this finding and the benefits of using online health-related information for diabetes information, physicians should promote the use of verified and credible diabetes websites, especially those with content in Arabic. Among those who used the Internet, there were only 71.6% (96/134) who searched for health-related information. Physicians may increase this number by promoting the credible and trustworthy websites to the 28.4% (38/134) of participants who use the Internet, but not for health-related information. Also, physicians should educate the more educated patients on how to search correctly and be able to critique health-related information using simple approaches. Because physicians are still the main and most trusted source of health information for most patients (216/344, 62.8%), involving physicians in the process of facilitating the online searching of diabetes information to patients is a strategy worth pursuing. With the positive impact it has on the health of patients, it may be a strategy that should be pursued by physicians.

Strengths and Limitations

Several studies have examined the general public's use of the Internet to obtain health-related information, but this study is the first research project to explore online health-related information–seeking behavior among Saudi adult patients diagnosed with type 2 diabetes and its impact on patients' self-care. This study's sample size is likely to be representative of the patient population in the primary care clinics and inpatients of one center, but may not be representative of the larger population. In this study, the majority of the participants were male. The main reason for this occurrence could be because the research team consisted of primarily males and females in Saudi Arabia often refuse to be interviewed by them due to cultural and social barriers. However, the response rate was high with only 9.4% (37/344) of patients who were approached declining to participate in the study. Because all participants were from a government hospital, most had relatively low education levels and low monthly income, were retired or unemployed, and the majority were married. The majority of the participants in this study did not receive any diabetic education by attending conferences or campaigns related to diabetes.

Conclusions

Among diabetes patients, less than half (134/344, 39.0%) of the sampled patients in this study used the Internet for health-related information. The majority of the participants who used the Internet belonged to the younger age group. Among Internet users, female participants were more likely to search for health-related information. The majority of the participants reported that their physician was the primary source of health-related information followed by television. Overall, this study demonstrates that those seeking health-related information are more aware about their health care needs as compared to non–health-related information seekers. These study results suggest that physicians should cooperate with their patients and guide them regarding reliable websites, which provide health information per patient need. To improve the health-related

information for patients, health care authorities should publish websites that contain reliable health information in the mother tongue so that patients learn better and are more aware of their health condition. The information must be updated and supervised regularly by health care providers. For those patients who do not use the Internet or cannot read, government organizations responsible for public health issues should make policies to reach them through alternative media, such as television, radio, and newspapers.

Acknowledgments

Conflicts of Interest

None declared.

References

1. Dutta-Bergman M. Trusted online sources of health information: differences in demographics, health beliefs, and health-information orientation. J Med Internet Res 2003;5(3):e21 [FREE Full text] [doi: 10.2196/jmir.5.3.e21] [Medline: 14517112]

2. AlGhamdi KM, Moussa NA. Internet use by the public to search for health-related information. Int J Med Inform 2012 Jun;81(6):363-373. [doi: 10.1016/j.ijmedinf.2011.12.004] [Medline: 22217800]

3. Eng T. The eHealth Landscape: A Terrain Map of Emerging Information and Communication Technologies in Health and Health Car. Princeton, NJ: The Robert Wood Johnson Foundation; 2001.

4. Weymann N, Dirmaier J, von WA, Kriston L, Härter M. Effectiveness of a Web-based tailored interactive health communication application for patients with type 2 diabetes or chronic low back pain: randomized controlled trial. J Med Internet Res 2015;17(3):e53 [FREE Full text] [doi: 10.2196/jmir.3904] [Medline: 25736340]

5. Nolan T, Dack C, Pal K, Ross J, Stevenson FA, Peacock R, et al. Patient reactions to a web-based cardiovascular risk calculator in type 2 diabetes: a qualitative study in primary care. Br J Gen Pract 2015 Mar;65(632):e152-e160 [FREE Full text] [doi: 10.3399/bjgp15X683953] [Medline: 25733436]

6. Moreau M, Gagnon M, Boudreau F. Development of a fully automated, web-based, tailored intervention promoting regular physical activity among insufficiently active adults with type 2 diabetes: integrating the I-change model, self-determination theory, and motivational interviewing components. JMIR Res Protoc 2015;4(1):e25 [FREE Full text] [doi: 10.2196/resprot.4099] [Medline: 25691346]

7. Glasgow RE, Kurz D, King D, Dickman JM, Faber AJ, Halterman E, et al. Twelve-month outcomes of an Internet-based diabetes self-management support program. Patient Educ Couns 2012 Apr;87(1):81-92 [FREE Full text] [doi: 10.1016/j.pec.2011.07.024] [Medline: 21924576]

8. de Boer MJ, Versteegen GJ, van Wijhe M. Patients' use of the Internet for pain-related medical information. Patient Educ Couns 2007 Sep;68(1):86-97. [doi: 10.1016/j.pec.2007.05.012] [Medline: 17590563]

9. Ferguson MO, Long JA, Zhu J, Small DS, Lawson B, Glick HA, et al. Low health literacy predicts misperceptions of diabetes control in patients with persistently elevated A1C. Diabetes Educ 2015 Jun;41(3):309-319. [doi: 10.1177/0145721715572446] [Medline: 25699568]

10. Hilliard ME, Sparling KM, Hitchcock J, Oser TK, Hood KK. The emerging diabetes online community. Curr Diabetes Rev 2015;11(4):261-272. [Medline: 25901500]

11. Guariguata L, Whiting DR, Hambleton I, Beagley J, Linnenkamp U, Shaw JE. Global estimates of diabetes prevalence for 2013 and projections for 2035. Diabetes Res Clin Pract 2014 Feb;103(2):137-149. [doi: 10.1016/j.diabres.2013.11.002] [Medline: 24630390]

12. Al-Nozha MA, Al-Maatouq MA, Al-Mazrou YY, Al-Harthi SS, Arafah MR, Khalil MZ, et al. Diabetes mellitus in Saudi Arabia. Saudi Med J 2004 Nov;25(11):1603-1610. [Medline: 15573186]

13. World Medical Association Declaration of Helsinki: Ethical Principles for Medical Research Involving Human Subjects. 2008. URL: http://www.wma.net/en/30publications/10policies/b3/17c.pdf [accessed 2015-08-07] [WebCite Cache ID 6abV91Pny]

14. Diaz JG, Ng J, Reinert S, Friedmann P, Moulton A. Patients use of the internet for medical information. J Gen Intern Med 2002;17(3):180-185. [doi: 10.1046/j.1525-1497.2002.10603.x]

15. El Bcheraoui C, Basulaiman M, Tuffaha M, Daoud F, Robinson M, Jaber S, et al. Status of the diabetes epidemic in the Kingdom of Saudi Arabia, 2013. Int J Public Health 2014 Dec;59(6):1011-1021. [doi: 10.1007/s00038-014-0612-4] [Medline: 25292457]

16. American DA. Standards of medical care in diabetes--2013. Diabetes Care 2013 Jan;36 Suppl 1:S11-S66 [FREE Full text] [doi: 10.2337/dc13-S011] [Medline: 25292457]

17. Siliquini R, Ceruti M, Lovato E, Bert F, Bruno S, De Vito E, et al. Surfing the internet for health information: an italian survey on use and population choices. BMC Med Inform Decis Mak 2011;11:21 [FREE Full text] [doi: 10.1186/1472-6947-11-21] [Medline: 21470435]

18. Ayers SL, Kronenfeld JJ. Chronic illness and health-seeking information on the Internet. Health (London) 2007 Jul;11(3):327-347. [doi: 10.1177/1363459307077547] [Medline: 17606698]

19. Grant RW, Cagliero E, Chueh HC, Meigs JB. Internet use among primary care patients with type 2 diabetes: the generation and education gap. J Gen Intern Med 2005 May;20(5):470-473 [FREE Full text] [doi: 10.1111/j.1525-1497.2005.04239.x] [Medline: 15963175]

20. Akerkar SK, Bichile L. Use of the Internet as a resource of health information by patients: a clinic-based study in the Indian population. J Postgrad Med 2005;51(2):116-118. [Medline: 16006703]

21. Holt K, Shehata A, Stromback J, Ljungberg E. Age and the effects of news media attention and social media use on political interest and participation: Do social media function as leveller? European Journal of Communication 2013 Feb 11;28(1):19-34. [doi: 10.1177/0267323112465369]

22. van Deursen AJ. Internet skill-related problems in accessing online health information. Int J Med Inform 2012 Jan;81(1):61-72. [doi: 10.1016/j.ijmedinf.2011.10.005] [Medline: 22079240]

23. Emont S. Measuring the Impact of Patient Portals: What the Literature Tells Us. Oakland, CA: California Health Care Foundation; 2011 May. URL: http://www.chcf.org/~/media/MEDIA%20LIBRARY%20Files/PDF/M/PDF%20MeasuringImpactPatientPortals.pdf [WebCite Cache ID 6aTgBXZDI]

24. Cline RJ, Haynes KM. Consumer health information seeking on the Internet: the state of the art. Health Educ Res 2001 Dec;16(6):671-692 [FREE Full text] [Medline: 11780707]

25. Thackeray R, Crookston B, West J. Correlates of health-related social media use among adults. JMIR 2013 Jan;15(1):000315113200026. [doi: 10.2196/jmir.2297]

26. Kruse RL, Koopman RJ, Wakefield BJ, Wakefield DS, Keplinger LE, Canfield SM, et al. Internet use by primary care patients: where is the digital divide? Fam Med 2012 May;44(5):342-347 [FREE Full text] [Medline: 23027117]

27. Longo D, Schubert S, Wright B, LeMaster J, Williams C, Clore J. Health information seeking, receipt, and use in diabetes self-management. Annals of Family Medicine 2010 Aug;8(4):334-340. [doi: 10.1370/afm.1115]

Abbreviations

IDF: International Diabetes Federation
MENA: Middle East and North Africa

Original Paper

"You Get Reminded You're a Sick Person": Personal Data Tracking and Patients With Multiple Chronic Conditions

Jessica S Ancker[1], MPH, PhD; Holly O Witteman[2,3,4], PhD; Baria Hafeez[1], MS; Thierry Provencher[5]; Mary Van de Graaf[1], MPH; Esther Wei[6], NP

[1]Weill Cornell Medical College, Department of Healthcare Policy and Research, Division of Health Informatics, New York, NY, United States

[2]Laval University, Faculty of Medicine, Department of Family and Emergency Medicine, Quebec City, QC, Canada

[3]Laval University, Faculty of Medicine, Office of Education and Continuing Professional Development, Quebec City, QC, Canada

[4]Research Centre of the CHU de Quebec, Quebec, QC, Canada

[5]Laval University, Quebec City, QC, Canada

[6]Weill Cornell Medical College, Department of Medicine, New York, NY, United States

Corresponding Author:
Jessica S Ancker, MPH, PhD
Weill Cornell Medical College
Department of Healthcare Policy and Research
Division of Health Informatics
425 E. 61st St., Suite 301
New York, NY, 10065
United States
Phone: 1 646 962 9404
Fax: 1 646 962 0105
Email: jsa7002@med.cornell.edu

Abstract

Background: Consumer health information technologies (HIT) that encourage self-tracking, such as diet and fitness tracking apps and disease journals, are attracting widespread interest among technology-oriented consumers (such as "quantified self" advocates), entrepreneurs, and the health care industry. Such electronic technologies could potentially benefit the growing population of patients with multiple chronic conditions (MCC). However, MCC is predominantly a condition of the elderly and disproportionately affects the less affluent, so it also seems possible that the barriers to use of consumer HIT would be particularly severe for this patient population.

Objective: Our aim was to explore the perspectives of individuals with MCC using a semistructured interview study. Our research questions were (1) How do individuals with MCC track their own health and medical data? and (2) How do patients and providers perceive and use patient-tracked data?

Methods: We used semistructured interviews with patients with multiple chronic diseases and providers with experience caring for such patients, as well as participation in a diabetes education group to triangulate emerging themes. Data were analyzed using grounded theory and thematic analysis. Recruitment and analysis took place iteratively until thematic saturation was reached.

Results: Interviews were conducted with 22 patients and 7 health care providers. The patients had an average of 3.5 chronic conditions, including type 2 diabetes, heart disease, chronic pain, and depression, and had regular relationships with an average of 5 providers. Four major themes arose from the interviews: (1) tracking this data feels like work for many patients, (2) personal medical data for individuals with chronic conditions are not simply objective facts, but instead provoke strong positive and negative emotions, value judgments, and diverse interpretations, (3) patients track for different purposes, ranging from sense-making to self-management to reporting to the doctor, and (4) patients often notice that physicians trust technologically measured data such as lab reports over patients' self-tracked data.

Conclusions: Developers of consumer health information technologies for data tracking (such as diet and exercise apps or blood glucose logs) often assume patients have unlimited enthusiasm for tracking their own health data via technology. However, our findings potentially explain relatively low adoption of consumer HIT, as they suggest that patients with multiple chronic illnesses consider it work to track their own data, that the data can be emotionally charged, and that they may perceive that providers do not welcome it. Similar themes have been found in some individual chronic diseases but appeared more complex because patients

XSL•FO

RenderX

often encountered "illness work" connected to multiple diseases simultaneously and frequently faced additional challenges from aging or difficult comorbidities such as chronic pain, depression, and anxiety. We suggest that to make a public health impact, consumer HIT developers should engage creatively with these pragmatic and emotional issues to reach an audience that is broader than technologically sophisticated early adopters. Novel technologies are likely to be successful only if they clearly reduce patient inconvenience and burden, helping them to accomplish their "illness work" more efficiently and effectively.

Introduction

Background

Consumer health information technology (HIT) is exploding in popularity, attracting the attention of technology-oriented consumers, patients, caregivers, and entrepreneurs. Technologies such as disease management apps and "quantified self" tools [1-3] offer the potential to help patients track personal data, learn about their health, and manage chronic care needs [4-7]. Consumer HIT appears poised to help inform, motivate, and engage patients, all of which are known to improve management skills and health outcomes [5-8].

However, it is not yet known whether such technologies will diffuse broadly beyond tech-savvy early adopters such as "quantified self" advocates, and whether the technologies would produce benefits for people with complex medical conditions. To date, the measured impact of consumer HIT is still limited. Computerized interventions for diabetes self-management have shown only limited efficacy [9,10]. In practice, effects have generally been limited as a result of low adoption and usage. One in 5 smartphone users has downloaded a health app [11], yet most apps are abandoned after a few uses [12]. Studies of the effectiveness of apps and websites to promote health outcomes (such as a recent study of a phone app to assist in weight loss [13] or a self-management Web community for diabetes [10]) frequently find that participants stop using the technology after a short period of time. Having a chronic condition increases the chances that a patient will use certain forms of consumer HIT on average [2,11]. But this increased likelihood is often offset by other sociodemographic factors that decrease the likelihood of using technology. Of particular concern from a public health standpoint, the use of consumer HIT remains lowest among the groups that might be most likely to benefit from additional forms of low-cost disease management support: people who are elderly, less educated, or less affluent [2,11,14]. These disparities in uptake, as well as the low rate of sustained use among adopters, suggest mismatches between current consumer HIT and the goals, desires, or capabilities of many patients [15,16].

A population with particularly complex and ongoing health needs is the 90 million Americans who have multiple chronic conditions (MCC) [17]. Although any combination of chronic conditions qualifies as MCC, the most common combinations are diabetes plus hypertension, heart disease plus hypertension, and cancer plus hypertension [18]. Patients with MCC experience the challenges associated with living with chronic disease and also typically consult more different doctors and coordinate more different therapeutic regimens than those with single diseases [17]. Each additional chronic condition places the individual at higher risk of adverse drug events, out-of-pocket expenses, impaired functional status, hospitalization, and mortality [17]. It is estimated that two-thirds of health care spending is focused on patients with MCC [17].

These patients are in need of improved strategies and technologies to support health and medical care, creating a number of opportunities that could potentially be filled with health IT, yet the barriers to technology adoption might be particularly problematic for these patients as well. MCC disproportionately affects the elderly and the less affluent. The prevalence of MCC rises sharply with age, affecting 34% of those aged 45-64 and 62% of those age 65 and over [19]. Furthermore, the prevalence of MCC is highest among the lowest income brackets, affecting nearly 51% of seniors who live at or below the federal poverty level but only 39% of seniors living at four times the poverty level [18].

As an initial step to exploring the perspectives of individuals with multiple chronic conditions, with the goal of understanding potential applications of consumer HIT and barriers to its use, we conducted a semistructured interview study. This paper focuses on tracking or keeping diaries of personal data, a task that we will refer to as "personal health information tracking". We focused on personal health information tracking because (1) it has been recommended for a variety of chronic conditions, and (2) it is a task potentially supported by consumer health IT. Self-monitoring tasks that have been promoted under different circumstances include blood glucose self-monitoring for certain patients with type 1 and type 2 diabetes [20,21], measuring blood pressure in hypertension and heart disease [22], keeping diet logs or food diaries for weight loss or digestive diseases [23], and self-monitoring medication adherence and side effects [24]. Patients also often receive the recommendation that they should check and be able to report certain laboratory values, such as CD4 count in human immunodeficiency virus (HIV) or hemoglobin A1c in diabetes. We therefore considered personal health information tracking to be a task that was likely to be encountered by patients with MCC, but we did not a priori assume a position on whether patients should self-track or whether it was likely to benefit them. Rather, our research questions were (1) How do individuals with MCC perform medical data tracking? and (2) How do patients and providers perceive and use patient-tracked data? We asked the questions broadly to encompass any sort of tool the patients were currently using, including electronic technologies, paper, or memory.

Theoretical Framework: Illness Work and Personal Health Information Management

This project was conducted from a human factors perspective influenced by the sociology of illness. This perspective recognizes that patients' management of their health comprises a wide variety of different activities both inside and outside the medical encounter: taking medicines, refilling prescriptions, buying and cooking food, exercising or doing physical therapy, researching health issues, coping with medical crises, finding doctors and dentists, organizing and traveling to medical appointments, and keeping records. As these are all effortful, directed activities to attain goals, they may be conceptualized as work [25-27].

Corbin and Strauss identified "illness work" as activities directly involved with managing an illness, such as following medication regimens and using technologies such as glucose meters or sleep apnea machines [25,26]. Yet even in illness, "everyday life work" of shopping, paying bills, nurturing relationships, and managing a household continues [25,26]. "Articulation work" is the planning, coordinating, and managing that allows people to complete all their other work [25,26].

Those components of illness and articulation work that involve acquiring and managing information can be called personal health information management [27-30]. A growing body of research on personal health information management has identified tasks including tracking health events, obtaining information, and organizing information [27]; creating personal histories, making decisions, planning, and structuring activities (eg, creating medication reminders) [28]; and transferring personal data and records to the physician [31]. In the current project, we focus on the subset of personal health information management involved in monitoring and logging personal data (such as symptoms or laboratory values), sometimes called personal health information tracking [32].

Much of the recent work in personal health information tracking and management has focused on generally healthy individuals and families [27,28,31,32], on patients with cancer [33-36], or (in support of information technology design) on computer-literate participants [31].

In this project, we sought to apply the insights from this previous work while exploring the perspectives of an economically diverse sample of patients with MCC in more depth. In order to develop or adapt technologies for these patients, it is essential to understand practices and perspectives of the potential users and the attributes of the tasks they seek to perform, as well as the social and physical environments in which they will be performing these tasks [15]. Poor fit between individuals, tasks, and technologies is likely to be one of the reasons that self-tracking technologies have not yet spread widely within populations with multiple diseases.

Methods

Participants

For individual interviews, we recruited purposive samples of adult English-speaking patients with MCC, and of medical providers with experience providing care for patients with MCC. We adopted the Department of Health and Human Services definition of chronic conditions as conditions that last a year or more and that either require ongoing medical attention or limit activities of daily living [17]. Patient participants were recruited from outpatient clinics in internal medicine and endocrinology and from the patient information library, using both promotional flyers and individual referrals from physicians and nurse practitioners. One researcher (JSA) also attended six 90-minute sessions of a diabetes education support group as a means of triangulating emerging themes. We chose the diabetes education group because many of the study participants had type 2 diabetes.

Settings

Weill Cornell Physicians is a multispecialty academic medical practice in Manhattan, with a mix of privately insured, Medicaid, and Medicare patients. New York-Presbyterian Hospital is the largest academic medical center in Manhattan. The Institute for Family Health is a federally qualified health center with 18 sites in and around New York City, providing safety net primary care to patients regardless of insurance status.

Interview Methods

The researchers developed a semistructured interview instrument centered on three topics: personal health information tracking, personal health information management, and searching for health-related information. The current manuscript focuses on the first of these. The first author conducted interviews in person, using offices and conference rooms convenient to the clinics where patients were recruited. Interviews were audio recorded and professionally transcribed. The interviewer also took field notes, collected samples of artifacts and documents for patients such as educational brochures, and photographed other artifacts or documents such as log sheets used to record glucose values.

Analysis Methods

No existing theoretical framework appeared to be appropriate to these data, and therefore we applied methods to develop meaning inductively from the data. Although this family of approaches is sometimes known in the sociology literature as development of grounded theory [37], we adopt the newer term "inductive thematic analysis" to reflect the fact that our end product is a series of interrelated themes rather than a fully formed theory [38]. Qualitative analysis was conducted collaboratively by our multidisciplinary team, which included individuals with training in journalism, public health, informatics, psychology, human factors, nursing, and diabetes education. Two of the researchers (HOW and EW) also brought personal experience of long-standing chronic disease. The preliminary version of the codebook was developed by 2 of the researchers in reading the first three transcripts and was iteratively refined over the coding process. Each transcript and photograph was reviewed by at least 2 team members (the first author and one other team member), who independently coded the transcript and then met to reach consensus on it.

We followed a staged and iterative approach, first identifying preliminary codes through repeated reading and review of the

data, then identifying relationships between codes and groupings of codes, and finally identifying and refining larger underlying themes. Over the analysis, 47 open codes were developed. These were linked into 6 broad groups: (1) resources, skills, and factors patients need for disease management, (2) the health care system and its components, (3) thinking, feeling, and experiencing disease and health, (4) medical data and medical records, (5) evaluative judgments, and (6) attributions of responsibility. In the final stage, the themes presented in the results section were developed.

To improve internal validity, we conducted member checking [39] in two ways. First, several of the emergent groups and themes were presented to new informants during interviews for their feedback. Second, the resulting themes were presented at a meeting of the diabetes education group, whose members validated the themes while also providing additional feedback and nuanced interpretation.

Analysis and recruitment were conducted simultaneously until saturation was achieved (ie, no new concepts were arising from new interviews) [40].

This study was approved by the Institutional Review Boards of Weill Cornell Medical College and the Institute for Family Health. All participants gave written informed consent. Members of the diabetes education group provided oral consent.

Results

Participants

Interviews were conducted with 22 patients and 7 health care providers. An additional 3 patient interviews were excluded from analysis because the interviewees did not have multiple chronic conditions.

The included patients reported having an average of 3.5 chronic conditions (SD 1.5). The most common conditions mentioned were type 2 diabetes, hypertension, heart disease, chronic pain, and depression. Other conditions included asthma, HIV, hepatitis C, thyroid disorders, rheumatoid arthritis, glaucoma, cataracts, and sleep apnea. Two individuals were in follow-up after cancer treatment. Conditions reported by only one patient each included type 1 diabetes, fibromyalgia, post-polio syndrome, sarcoidosis,

Sjogren syndrome, and cirrhosis. Many described themselves as overweight but none as obese. In addition to their chronic conditions, patients also discussed a wide variety of recently experienced urgent conditions, including diverticulitis, flu, appendicitis, bee stings, and physical injuries. Participants sometimes mentioned taking drugs that implied other chronic conditions that they did not explicitly list: examples included antidepressants, blood pressure medications, lipid-lowering medications, drugs for prostatic hyperplasia, and anticoagulants. Many of the patients with type 2 diabetes were taking insulin one or more times a day, as was the individual with type 1 diabetes.

Half of patients were men and half were women; a third (n=7) were black. Ages ranged from 37-89 (mean 64.1; median 66). About two-thirds (n=15) were not currently married. Just over a third (n=8) used English as a second language. One third (n=7) were covered by Medicare (US public insurance for those over age 65); one third (n=7) by Medicaid (US public insurance for those with low income); and the remainder (n=8) by commercial insurance.

Multiple chronic conditions placed heavy and sometimes competing demands on patients. For example, one patient with diabetes recognized that his morning toast caused increases in his blood glucose, but on balance had decided not to stop eating toast because his morning medications for other conditions had to be taken with food. Several patients with diabetes or heart disease recognized that exercise might help but were prevented because of chronic pain or disability from injury. Patients taking anticoagulants encountered challenges when scheduling surgery for other conditions.

The diabetes education group was attended by an average of 5 patients each session (range 4-9). Most patient education group attendees had type 2 diabetes but a minority had type 1 diabetes or prediabetes.

The health care providers were 2 nurse practitioners, 2 internists, 2 family medicine physicians, and an emergency medicine physician.

Major themes pertaining to personal health information tracking are summarized in Table 1 and presented in detail in the results section.

Table 1. Major themes in personal health information tracking.

Themes	Summary	Representative quotes
1. Personal data can carry strong emotional and moral implications	Data are not merely objective facts but prompt strong positive and negative emotions as well as value judgments.	"You get reminded you're a sick person" and "I'm not a good patient".
2. Multiple purposes and uses for personal data	Patients use data for a variety of purposes, ranging from active self-management to making sense of their condition to reporting to the doctor.	"I'll [check] it if I'm feeling lightheaded".
3. (Un)reliability of personally tracked data	Patients often notice that physicians do not trust their self-tracked data.	"[The doctors] looked at [my logs] very superficially…they seem to rely on your A1c numbers".
4. Tracking feels like work	Tracking is time-consuming and sometimes emotionally draining.	"It's too cumbersome for me".

Overview

Most patients paid attention to laboratory findings provided by their doctors, and a few kept records of selected values. For example, a woman with anemia created a table to track her blood test results over time, 2 patients with HIV kept records of their CD4 count values over time, and many patients checked on their cholesterol regularly.

However, fewer than half regularly tracked data by self-testing or recording daily activities. The most common example of self-tracking was patients with diabetes monitoring their blood glucose. Among the 16 patients with type 1 or type 2 diabetes, 11 mentioned self-monitoring blood glucose in some fashion (some were fairly regular, some checked values occasionally, and some said they used to monitor regularly but had stopped). Other examples of tracking mentioned by one or more participants included recording weight or blood pressure (n=7), tracking daily medication administration (n=3), keeping food diaries (n=2, in one case to investigate suspected lactose intolerance), collecting laboratory reports to manually compare trends over time (n=4), and recording potential side effects with a new medication (n=2). This sort of tracking was conducted on paper or electronically on a spreadsheet, or in one case on a paper calendar. All the patients interviewed who monitored blood glucose used monitors that tracked data electronically. In addition, some kept handwritten blood glucose logs. The numbers in parentheses above (n=) are provided for perspective, but these data were collected through open-ended interview questions rather than closed-ended survey methods, so the interviews may not have captured every instance of tracking.

Many of the patients older than 65 and most Medicaid patients did not use computers regularly or at all, and many did not have smartphones.

Theme 1: Personal Data Can Carry Strong Emotional and Moral Implications

Overview

Indicators such as blood glucose, weight, and lab values were not discussed as value-free facts but instead carried strong emotional and evaluative connotations. People recognized tracking as work, judged themselves as "good" or "bad" for their data and their diligence in collecting it, and noted that data should be considered within the patient's personal context.

Negative Aspects of Illness

Medical data often reminded patients of the negative aspects of their illness. An individual who did not monitor her blood glucose regularly said her values were "depressing", and another said they made her "scared". Discussing tracking sometimes raised feelings of anger or injustice not only about the tracking but also about having chronic disease. "I hate to be focused on my health in every friggin' second of the day...I don't want to live like that every day". A patient with HIV, hypertension, and other chronic illnesses said he avoided looking at his regular test results: "I don't ask about no numbers. If anything is messing up, then [my doctor] tells me". The physical experience could also be unpleasant. "Poking my finger, that was irritating to me," said one person who had abandoned blood glucose

self-monitoring. "I'm tired of sticking myself," another said. Some patients with diabetes said they were frustrated to see their blood glucose values occasionally spike without a clear reason, undermining their confidence that they understood and could manage their disease.

The Moral Valence of Medical Data

Patients and providers frequently described the data with highly judgmental language, including terms suggesting moral transgression. For example, one explained a high blood glucose value because "I cheated and I had some McDonald's". Conversely, patients could feel extremely happy and proud when their values were good. Several of the health care providers said it was better to use nonjudgmental language such as "high/low" or "target/nontarget" because patients "get discouraged because they think they're being graded or judged". Yet in the interviews, many providers used more evaluative language such as "good/bad" and "better/worse". A patient who had altered his diet and was able to lower his doses of hypertension and hyperlipidemia drugs said he felt satisfied when his doctors said, "Okay, we're happy with you".

The Moral Valence of Tracking

There was also a "good/bad patient" aspect to tracking itself. People with diabetes frequently called themselves a "bad patient" or "not a good patient" when they did not monitor blood glucose. One participant explained the fact that she did not track any of her health indicators (including diet and exercise) by calling herself "lazy". Although providers most often expressed frustration about lack of monitoring, some occasionally perceived monitoring as excessive. Patients who tracked data very diligently (eg, detailed exercise logs, which clinicians saw as having little clinical relevance) were sometimes referred to as "obsessive and compulsive" or "fastidious".

My Interpretation of My Data

Although in some cases patients and physicians were in close agreement about what data values were "good" or "bad", other patients preferred to interpret their results in light of their own unique histories or symptoms. For example, several patients with diabetes said that they aimed for a blood glucose level or hemoglobin A1c that was appropriate "for me". In some cases, these were values that made them feel well, or values that were high enough to minimize the risk of hypoglycemia. In other cases, patients wanted their personal history to be taken into account in interpreting data. For example, a person with a history of obesity took pride in the number of dress sizes she had gone down, rather than aiming for a particular target weight. One provider told an anecdote about a patient who had brought her hemoglobin A1c from 13% to below 8% with diet and medication. When urged to continue lowering it, the patient said, "I don't want to be a poster child for perfect diabetes". The doctor recalled saying, "Actually, you're right. This is good for you...I should've been jumping up and down because that's really great".

Theme 2: Multiple Purposes and Uses for Personal Data

Overview

Not all patients closely monitored their own data values. Patients who did track their own data through either self-monitoring or laboratory testing described a variety of purposes, which depended on aspects of their disease and on their own experience of their disease. They might use their tracked data for real-time decision making, for medium-term self-assessment, or for making sense of various elements of data, such as physical symptoms.

Tracking for Action

Some experienced patients with diabetes monitored blood glucose multiple times per day as "working data" [30] that they would use immediately to adjust their diet or their medication. For example, one woman described a highly effective routine of using thrice-daily glucose monitoring to adjust sliding-scale medication doses and diet. She had used these techniques to reduce her hemoglobin A1c level to 6.1% for nearly a year. Most health care providers perceived this active, real-time use of data for self-management as important for patients who were struggling to manage conditions in which data values were highly sensitive to behavior (such as a younger patient with new-onset diabetes), but less important for others (such as older patients with stable disease).

Tracking for Goal-Checking

A second approach was to use data periodically to assess progress toward a goal. Patients with this approach referred to the data for a holistic assessment of how "well" they were doing, but not necessarily for active, hour-to-hour self-management. This was also often the approach used by patients who were monitoring indicators that they themselves could not measure, such as cholesterol, blood count values in anemia, HIV viral load levels, and CD4 counts.

Tracking for Sense-Making

A different approach was to examine data values as part of trying to make sense of the disease. Several patients with diabetes who did not regularly monitor described checking glucose when they felt symptoms they suspected indicated hypoglycemia: "I'll do it if I'm feeling lightheaded". Another said he did it when he felt a "hunch". This approach was sometimes encouraged by physicians for patients who seemed unlikely to monitor regularly: "Usually I tell them that if they're not feeling well, check their blood sugar". One patient with HIV asked his doctor for explanations whenever his lab values changed. "I saw this is different [from] last 2-3 months ago, and now something is wrong. And he explained to me if it's something wrong or not [important]". During visits, health care providers frequently explicitly linked lab values to patient behavior to encourage them to develop a more biomedical concept of the disease. For example, one provider used a patient's headache as a teaching example to discuss the role of salt in her diet. Some also saw it as a useful short-term exercise for patients seeking an understanding of behavioral triggers for conditions such as asthma, irritable bowel syndrome, or migraine

headaches. However, some patients described frustration (or even abandoning tracking altogether) after failing to see connections between their data values and their behavior.

Tracking for the Doctor

A few patients appeared to perceive self-monitoring as something done not for their own use but partly or largely to create records for the doctor. A few seemed confused that doctors rarely reviewed their logs. "They don't monitor that part of it, I don't know why".

Theme 3: (Un)reliability of Personally Tracked Data

Providers often perceived patient-recorded data as unreliable. The lack of confidence was attributed to perceived lack of diligence, moral valence of the data (with patients unwilling to "admit" undesirable numbers), and fear of consequences. The most striking example, told by a provider, was a woman who faked her daughter's blood glucose log to persuade the doctor to delay starting insulin therapy.

Providers sometimes described lab data as more trustworthy than data from self-tracking. "The hemoglobin A1c don't lie [sic], so you can tell me whatever you want, but it's going to tell me the truth of what's going on in your body". Another said: "For the most part a lot of this information I don't really [need] because I can check the A1c and know what it's like". Current diabetes treatment guidelines recommend attention to self-monitored blood glucose for extreme values and trends, in addition to hemoglobin A1c as an indicator of overall control [20,21].

These perceptions on the part of providers were evident to many of the patients. "I remember when I used to go to the diabetes center up there with [a doctor] and she looked at it very superficially too, and they seem to rely on your A1c numbers," said a patient who had abandoned logging his daily glucose values. Providers also sometimes perceived automated recording devices as more reliable than patient-recorded information, which was also noticed by some of the patients: "[My doctor] is like, 'Please bring me the machine'". One provider told an anecdote about a patient with a dangerous blood pressure increase; the patient's spouse used a monitor to print out the previous week's blood pressure readings, which were low enough to persuade the doctors to rule out their initial suspicion of "medication noncompliance".

In only one case, a highly engaged patient said that her provider preferred reviewing her blood glucose logs rather than the glucose monitor because the log made it easier to link the readings to meals. "It was a lot of confusion with the doctor because I was just bringing the machine. So now [with the notebook] they know that first one, two, three is breakfast, lunch, and dinner".

Theme 4: Tracking as Work

Patients said that tracking was effortful and time-consuming, sometimes explicitly describing it as work. A patient with diabetes said it was a waste of time to write down her values: "I'm not going to sit down and write a paper for the month to keep track of it". One woman noted that she kept medical information about her multiple conditions, as well as her

multiple health care providers, in her office rather than her home. Data tracking sometime was felt to conflict with the work of everyday living forcing trade-offs when patients did not have sufficient time or emotional resources. A diabetes patient who had given up self-monitoring of blood glucose said, "It's too cumbersome for me". A patient with heart disease who kept a diet log gave it up after it became "overwhelming".

Discussion

Principal Findings

Developers of consumer health information technologies for data tracking (such as diet and exercise apps or blood glucose logs) often assume patients have unlimited enthusiasm for tracking their own health data via technology, that these data are objective facts with unambiguous interpretations and applications, and that health care providers welcome such data in their assessment of a patient's health status. Potential users are believed to be "willing to assume a more participatory role in the management of their health, to learn how to use new tools, and to commit themselves to doing so constantly" [31].

By contrast, the concept of data tracking as patient work was strongly supported by our interviews with patients with multiple chronic conditions. Furthermore, personal medical data did not appear to be objective facts, interpreted in the same way by patients and their providers. The data provoke strong negative and positive emotional reactions, sometimes overwhelming ones that prevent people from wanting to track or access their data. These data can also make individuals feel judged by their health care providers or even by themselves. Patients may resist their physician's interpretation of their data values as "one-size-fits-all" and may prefer to weight their own personal history and disease experience. Physicians often trust technologically measured data more than manual self-tracked data; their preference is apparent to patients and may inadvertently be sending patients mixed messages about the value of their data tracking efforts.

Our study also suggests that patients who do keep track of their data require it for different purposes. Some patients examine their data periodically for a holistic check on their own progress toward goals, and others use their data for real-time decisions about their behavior. Yet another group of individuals inspect and interpret this data as part of the process of developing an understanding of their disease.

Finally, we encountered many elderly and low-income patients who had limited experience with and access to electronic technologies. As our sample was fairly representative of the demographics of those with MCC (with a mean age of 64 and about one third covered by Medicaid), it is plausible that this reflects the experience of broader MCC populations.

These findings support the proposal that existing self-tracking technologies such as mobile phone apps may not provide a good fit to the needs and abilities of individuals with MCC and the tasks they are seeking to perform with them [15].

Comparison With Prior Work

Our work contributes to a growing body of research in personal health information management and personal health information tracking—research that has already identified a range of tasks frequently performed by patients, ranging from tracking health data to managing medical records to creating personal reminder systems [27-29,31,32]. However, much of the previous work in this field has focused on generally healthy individuals and families [27,28,31,32], on patients with cancer [33-36], or (in support of HIT design) on computer-literate participants [31].

The current project identifies different perspectives brought by an economically diverse group of patients with multiple chronic diseases. Our participants each had several chronic diseases, including diabetes, HIV, heart disease, depression, and many others, and about one third were covered by Medicaid. Their perspectives were in many cases different from what has been found in previous work with healthy families. For example, while healthy consumers in Canada rejected the idea that health information management was "work" [32], our patients with MCC frequently described managing data as time-consuming and tiring. There are several potential explanations for this contrast. First, keeping track of even a single chronic disease is likely to be more challenging than keeping track of preventive care or minor medical events among largely healthy individuals. Second, individuals with multiple chronic conditions are likely to have "illness work" connected to each of the diseases (our patients had an average of 3.5 chronic conditions). Third, MCC is disproportionately a condition of the elderly as well as the less affluent, meaning that an MCC patient may be conducting "illness work" while simultaneously facing challenges related to aging and poverty. Finally, the multiple chronic conditions included physically and emotionally challenging comorbidities such as depression, anxiety, and chronic pain—conditions that themselves might make it more difficult to conduct any "illness work". This workload burden may have been particularly evident as many of our patients were unmarried and had primary responsibility for their own personal health information. By contrast, previous research with families often shows that one family member takes primary responsibility for the information needs of the household [27,28,31,32]. Such a division of labor within the family context might offer several advantages, including the ability for the information manager to specialize and develop expertise in information management, and might also alleviate the workload burden on more ill members of the household.

Our findings support previous work in the field of technology development for elderly patients or others who do not use electronic technologies regularly. The people we spoke with conducted personal health information management and tracking with a variety of paper and electronic tools, both custom-made and adapted, as has been found by other researchers [27,29,31]. As others have found, we found that older patients and those with Medicaid were frequently unfamiliar with electronic technologies. In addition to lack of access, some have found that elderly patients may find usability barriers discouraging them from adopting new technologies [41]. We additionally found that some adults with experience of chronic disease have already solved their own data management problems to their

own satisfaction and did not express much interest in novel technologies. Similarly, Grindrod et al found that older patients, when introduced to new technologies, "struggled to think of a need for the applications in their own lives" [41].

When combined with our finding that patients considered data tracking to be "cumbersome", this suggests that novel technologies will succeed only if they are highly intuitive, easy to learn, and unambiguously reduce the burden of work on the patient. Uploadable device data [42] or mining of personal data traces from phones and other technologies [43,44] may be effective ways of accomplishing this, especially given the fact that both patients and providers in our study recognized the additional perceived credibility of technologically measured data. The gamification trend in the health promotion and disease management literature is also potentially relevant [42,45,46]. Games that provide motivation to track learning opportunities, social support, or emotional coping support for dealing with data could potentially be useful for patients with chronic disease. However, designers of games for self-tracking may wish to consider our findings that patients often see data tracking as work and may perceive the data as having moral meaning that could be positive or negative. As noted by others, patients can have strong emotional responses to learning their own numbers and can feel judged by themselves and others [47-49]. Turning information tracking into a game might appear to trivialize important tasks, and "losing" in a game might amplify negative emotions. It might even be that some patients might prefer less emotionally charged technologies inspired by office or financial management software, which are explicitly designed to make necessary activities efficient and even pleasant while still treating those activities as work.

Our findings also have relevance for the literature on patients' mental models of disease. As others [50,51] have pointed out, individuals work to make sense of their disease and health experiences, seeking a label or name, identifying its cause, establishing its probable timeline and consequences, and learning the extent to which it is manageable or curable. Over time, people use these insights to construct what have been called "common-sense models of disease" or "illness representations", that is, explanations of health conditions that are internally coherent but that may or may not coincide with the biomedical model of the disease [50,51]. These illness representations can affect risk perceptions, coping behavior, management, and disease outcomes. Data tracking clearly offers the possibility of demonstrating the link between behavior and disease indicators (eg, between diet, medication administration, and blood glucose), thereby encouraging patients to develop a more biomedical model of their disease.

However, not all patients wanted to examine their data for this purpose. Our findings are striking in the degree to which medical data were shown to have extremely serious emotional implications for patients with MCC, sometimes serious enough to be associated with abandonment of data tracking altogether. "Bad" data values can be extremely upsetting, especially when those "bad" values have or are perceived to have some link to behavior. Patients' language revealed the extent to which they use judgmental terms of sin and transgression to describe both their data and themselves. Furthermore, some of our patients

noted with some surprise that their health care providers did not seem very interested in their self-logged data; others have noted that diabetes patients can interpret their providers' preference for lab-measured hemoglobin A1c as meaning that self-monitoring was not important [48]. Peel et al found that counterintuitive blood glucose values confused patients and could lead to discontinuation of monitoring, as was reported by one of our patients [49].

One highly relevant study reports a trial of an electronic diabetes diary and information app, which incorporates some of the concepts we have recommended here [52]. In that trial, blood glucose measurements were automatically uploaded via Bluetooth from an electronic monitor, although food and exercise data had to be manually input. Counseling, including motivational interviewing, was added in one of the two technology arms. Nevertheless, after 4 months the app (with or without supplemental counseling) was not associated with changes in hemoglobin A1c levels [52]. The 18% attrition rate in this study may have resulted from the relatively heavy work burden of self-tracking the electronic data.

Limitations

The sample was generally representative of the demographics of the MCC population. However, type 2 diabetes may have been more prevalent in our sample than in the national MCC population, in which type 2 diabetes occurs in three of the top nine pairs of chronic conditions and four of the top nine condition triads [53]. Interviews were also conducted in a US urban area and in English only, limiting the sample to patients comfortable in that language. These reasons may limit relevance to other populations, such as individuals in other countries with different health care systems, people in rural locations with different challenges in accessing health care, or people of other cultures or language groups.

Conclusions and Implications

Developers of consumer health information technologies for data tracking (such as diet and exercise apps or blood glucose logs) often assume that a wide variety of patients will have unlimited enthusiasm for tracking their own health data via technology. However, adoption of new technologies does not always rapidly spread beyond computer-literate, highly motivated early adopters. We suggest that to make a public health impact, developers should be prepared to engage creatively with a variety of pragmatic and emotional issues to reach a broader audience that includes patients with chronic disease.

One recommendation is to explore ways to engage directly with the emotional impact associated with medical data, exploring ways not only to motivate progress but also cope with negative feelings. Developers should seek not to exacerbate negative feelings or judgments, look at creative ways to support positive feelings, and facilitate personal goal setting rather than imposing external goals. Technologies could integrate techniques such as motivational interviewing [54] that have been demonstrated to help patients establish personally relevant goals and action plans, rather than seeking to persuade patients to adopt their providers' priorities. The behavioral economics literature can

provide valuable guidance in leveraging effects such as framing, defaults, and behavioral "nudges" to promote engagement and better decision making [55].

Another suggestion is to provide different formats for different purposes. Patients who are building a conceptual understanding of disease might benefit from data-driven links with explanatory material or even simulations. Patients who are using data to check on goals might benefit from progress bars or visualized target thresholds. A relatively small number of patients (such as those adjusting insulin doses or high blood pressure medications [22]) will be using data for self-management; these individuals are most likely to be interested in reminders or alerts. Developing systems with the wrong purpose in mind appears likely to irritate patients rather than support them. For example, patients who have not established personally relevant goals are unlikely to welcome visualizations that depict their "progress", and patients who are already well educated about their disease processes may prefer emotional and practical support to basic educational material.

It must still be recognized that older generations are not universally comfortable with electronic technologies and that many low-income patients still do not have access to them. For the foreseeable future, a significant subset of patients will lack access to information technology. This creates tremendous opportunities for exploring improved paper technologies. For example, scannable paper forms might ease the burden of tracking data on paper and be more widely used than mobile apps by some groups. Technologies that benefit only younger or more technologically sophisticated patients could have the potential to widen health disparities rather than narrow them. This issue of equity must be addressed in health information technology broadly, but especially in technology intended for personal health information tracking and management.

Finally, the concept of data tracking as yet another piece of patient "work" resonated strongly with the participants. Novel technologies are likely to be successful only if they clearly reduce inconveniences and burden for patients, helping them to accomplish their "work" more efficiently and effectively.

Acknowledgments

References

1. Swan M. Emerging patient-driven health care models: an examination of health social networks, consumer personalized medicine and quantified self-tracking. Int J Environ Res Public Health 2009 Feb;6(2):492-525 [FREE Full text] [doi: 10.3390/ijerph6020492] [Medline: 19440396]

2. Ancker JS, Barrón Y, Rockoff ML, Hauser D, Pichardo M, Szerencsy A, et al. Use of an electronic patient portal among disadvantaged populations. J Gen Intern Med 2011 Oct;26(10):1117-1123 [FREE Full text] [doi: 10.1007/s11606-011-1749-y] [Medline: 21647748]

3. Phelps RG, Taylor J, Simpson K, Samuel J, Turner AN. Patients' continuing use of an online health record: a quantitative evaluation of 14,000 patient years of access data. J Med Internet Res 2014;16(10):e241 [FREE Full text] [doi: 10.2196/jmir.3371] [Medline: 25344621]

4. Gibbons C, Wilson R, Samal L, Lehmann C, Dickersin K, Aboumatar H. Impact of consumer health informatics applications. Bethesda, MD: Agency for Healthcare Research and Quality, US Department of Health and Human Services,; 2009.

5. Greenhalgh T. Patient and public involvement in chronic illness: beyond the expert patient. BMJ 2009 Feb 17;338(feb17 1):b49-b49. [doi: 10.1136/bmj.b49]

6. Hibbard JH, Greene J. What the evidence shows about patient activation: better health outcomes and care experiences; fewer data on costs. Health Aff (Millwood) 2013 Feb;32(2):207-214. [doi: 10.1377/hlthaff.2012.1061] [Medline: 23381511]

7. Hibbard JH, Mahoney ER, Stock R, Tusler M. Do increases in patient activation result in improved self-management behaviors? Health Serv Res 2007 Aug;42(4):1443-1463 [FREE Full text] [doi: 10.1111/j.1475-6773.2006.00669.x] [Medline: 17610432]

8. Nolte S, Elsworth GR, Sinclair AJ, Osborne RH. The extent and breadth of benefits from participating in chronic disease self-management courses: a national patient-reported outcomes survey. Patient Educ Couns 2007 Mar;65(3):351-360. [doi: 10.1016/j.pec.2006.08.016] [Medline: 17027221]

9. Pal K, Eastwood S, Michie S, Farmer A, Barnard M, Peacock R. Computer-based diabetes self-management interventions for adults with type 2 diabetes mellitus. Cochrane Database Syst Rev 2013;3:1-100. [doi: 10.1002/14651858.CD008776.pub2]

10. Yu CH, Parsons JA, Mamdani M, Lebovic G, Hall S, Newton D, et al. A web-based intervention to support self-management of patients with type 2 diabetes mellitus: effect on self-efficacy, self-care and diabetes distress. BMC Med Inform Decis Mak 2014;14:117 [FREE Full text] [doi: 10.1186/s12911-014-0117-3] [Medline: 25495847]

11. Fox S, Duggan M. Pew Research Center's Internet & American Life Project. 2013. Mobile Health 2012 URL: http://www.pewinternet.org/files/old-media//Files/Reports/2012/PIP_MobileHealth2012_FINAL.pdf [accessed 2015-08-12] [WebCite Cache ID 6ajWW0fHr]

12. App retention improves: Apps used only once declines to 20%. 2014. URL: http://info.localytics.com/blog/app-retention-improves [accessed 2015-08-12] [WebCite Cache ID 6ajWYclWe]

XSL•FO
RenderX

13.   Laing BY, Mangione CM, Tseng C, Leng M, Vaisberg E, Mahida M, et al. Effectiveness of a smartphone application for weight loss compared with usual care in overweight primary care patients: a randomized, controlled trial. Ann Intern Med 2014 Nov 18;161(10 Suppl):S5-12 [FREE Full text] [doi: 10.7326/M13-3005] [Medline: 25402403]

14.   Roblin DW, Houston TK, Allison JJ, Joski PJ, Becker ER. Disparities in use of a personal health record in a managed care organization. J Am Med Inform Assoc 2009;16(5):683-689 [FREE Full text] [doi: 10.1197/jamia.M3169] [Medline: 19567790]

15.   Ammenwerth E, Iller C, Mahler C. IT-adoption and the interaction of task, technology and individuals: a fit framework and a case study. BMC Med Inform Decis Mak 2006;6:3 [FREE Full text] [doi: 10.1186/1472-6947-6-3] [Medline: 16401336]

16.   Ozkaynak M, Brennan P, Hanauer D, Johnson S, Aarts J, Zheng K. Patient-centered care requires a patient-oriented workflow model. Journal of the American Medical Informatics Association 2013;20:e14-e16. [doi: 10.1136/amiajnl-2013-001633]

17.   US Department of HealthHuman Services. Multiple Chronic Medical Conditions and Health-Related Quality of Life in Older Adults, 2004–2006. Preventing Chronic Disease 2010;10:120282-120282. [doi: 10.5888/pcd10.120282]

18.   Fried V, Bernstein A, Bush M. Multiple chronic conditions among adults aged 45 and over: Trends over the past 10 years. NCHS Data Brief. Hyattsville, MD: National Center for Health Statistics; 2012.

19.   Ward BW, Schiller JS, Goodman RA. Multiple chronic conditions among US adults: a 2012 update. Prev Chronic Dis 2014;11:E62 [FREE Full text] [doi: 10.5888/pcd11.130389] [Medline: 24742395]

20.   American Association of Diabetes Educators. Diabetes Educator. 2010. Self-monitoring of blood glucose using glucose meters in the management of type 2 diabetes URL: http://www.diabeteseducator.org/export/sites/aade/_resources/pdf/research/SelfMonitoring2010.pdf [accessed 2015-01-05] [WebCite Cache ID 6VO5IH5El]

21.   American Diabetes Association. Standard of Medical Care in Diabetes 2014. Diabetes Care 2014;37(S1):S14-S80. [doi: 10.2337/dc14-S014]

22.   McManus RJ, Mant J, Haque MS, Bray EP, Bryan S, Greenfield SM, Hobbs F D Richard, et al. Effect of self-monitoring and medication self-titration on systolic blood pressure in hypertensive patients at high risk of cardiovascular disease: the TASMIN-SR randomized clinical trial. JAMA 2014 Aug 27;312(8):799-808. [doi: 10.1001/jama.2014.10057] [Medline: 25157723]

23.   Pignone MP, Ammerman A, Fernandez L, Orleans CT, Pender N, Woolf S, et al. Counseling to promote a healthy diet in adults: a summary of the evidence for the U.S. Preventive Services Task Force. Am J Prev Med 2003 Jan;24(1):75-92. [Medline: 12554027]

24.   Zullig LL, Peterson ED, Bosworth HB. Ingredients of successful interventions to improve medication adherence. JAMA 2013 Dec 25;310(24):2611-2612. [doi: 10.1001/jama.2013.282818] [Medline: 24264605]

25.   Corbin J, Strauss A. Managing chronic illness at home: Three lines of work. Qualitative Sociology 1985;8(3):224-247.

26.   Valdez R, Holden R, Novak L, Veinot T. Transforming consumer health informatics through a patient work framework: Connecting patients to context. Journal of the American Medical Informatics Association 2014:2-10. [doi: 10.1136/amiajnl-2014-002826]

27.   Moen A, Brennan PF. Health@Home: the work of health information management in the household (HIMH): implications for consumer health informatics (CHI) innovations. J Am Med Inform Assoc 2005;12(6):648-656 [FREE Full text] [doi: 10.1197/jamia.M1758] [Medline: 16049230]

28.   Civan A, Skeels M, Stolyar A, Pratt W. Personal health information management: consumers' perspectives. AMIA Annual Symposium Proceedings 2006:156-160 [FREE Full text] [Medline: 17238322]

29.   Zayas-Cabán T. Health information management in the home: a human factors assessment. Work 2012;41(3):315-328. [doi: 10.3233/WOR-2012-1306] [Medline: 22398501]

30.   Agarwal R, Khuntia J. Personal health information in the design of consumer health information technology: Background report. In: AHRQ Publication number 09-0075-EF. Rockville, MD: Agency for Healthcare Research and Quality; 2009.

31.   Piras E, Zanutto A. Prescriptions, X-rays and Grocery Lists. Designing a Personal Health Record to Support (The Invisible Work Of) Health Information Management in the Household. Comput Supported Coop Work 2010 Nov 3;19(6):585-613. [doi: 10.1007/s10606-010-9128-5]

32.   MacGregor JCD, Wathen CN. 'My health is not a job': a qualitative exploration of personal health management and imperatives of the 'new public health'. BMC Public Health 2014;14:726 [FREE Full text] [doi: 10.1186/1471-2458-14-726] [Medline: 25030501]

33.   Unruh KT, Pratt W. The Invisible Work of Being a Patient and Implications for Health Care: "[the doctor is] my business partner in the most important business in my life, staying alive". 2008 Presented at: Ethnographic Praxis in Industry Conference; 2008; Prague, Czech Republic p. 40-50. [doi: 10.1111/j.1559-8918.2008.tb00093.x]

34.   Unruh KT, Pratt W. Barriers to organizing information during cancer care: "I don't know how people do it.". AMIA Annu Symp Proc 2008:742-746 [FREE Full text] [Medline: 18999223]

35.   Klasnja P, Hartzler AC, Unruh KT, Pratt W. Blowing in the Wind: Unanchored Patient Information Work during Cancer Care. In: CHI Conf Proc. 2010 Presented at: Conference on Human Factors in Computing Systems; 2010; Atlanta, GA p. 193-202 URL: http://europepmc.org/abstract/MED/20697453 [doi: 10.1145/1753326.1753355]

36. Unruh KT, Skeels M, Civan-Hartzler A, Pratt W. Transforming Clinic Environments into Information Workspaces for Patients. In: CHI Conf Proc. 2010 Presented at: Conference on Human Factors in Computing Systems; 2010; Atlanta, GA p. 183-192 URL: http://europepmc.org/abstract/MED/21654895 [doi: 10.1145/1753326.1753354]

37. Strauss A, Corbin J. Basics of Qualitative Research: Techniques and Procedures for Developing Grounded Theory. 2nd ed. Thousand Oaks, CA: Sage; 1998.

38. Braun V, Clarke V. Using thematic analysis in psychology. Qualitative Research in Psychology 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]

39. Morse JM, Barrett M, Mayan M, Olson K, Speirs J. Verification strategies for establishing reliability and validity in qualitative research. International Journal of Qualitative Methods 2002;1(2):1-19.

40. Morse J. The Significance of Saturation. Qualitative Health Research 1995 May 01;5(2):147-149. [doi: 10.1177/104973239500500201]

41. Grindrod KA, Li M, Gates A. Evaluating user perceptions of mobile medication management applications with older adults: a usability study. JMIR Mhealth Uhealth 2014;2(1):e11 [FREE Full text] [doi: 10.2196/mhealth.3048] [Medline: 25099993]

42. Grifantini K. Advances in management technology for diabetes. IEEE Pulse 2013:40-44.

43. Estrin D. Small data, where n=me. Communications of the ACM 2014;57(4):32-34.

44. Hsieh C, Tangmunarunkit H, Alquaddoomi F, Jenkins J, Kang J, Ketcham C. Lifestreams: A modular sense-making toolset for identifying important patterns from everyday life. In: Sensys '13: Proceedings of the 11th ACM Conference on Embedded Networked Sensor Systems. 2013 Presented at: ACM Conference on Embedded Networked Sensor Systems; Nov. 11-14, 2013; Rome, Italy. [doi: 10.1145/2517351.2517368]

45. Leahey T, Rosen J. DietBet: A Web-Based Program that Uses Social Gaming and Financial Incentives to Promote Weight Loss. JMIR Serious Games 2014;2(1):e2 [FREE Full text] [doi: 10.2196/games.2987] [Medline: 25658966]

46. Ruggiero L, Moadsiri A, Quinn LT, Riley BB, Danielson KK, Monahan C, et al. Diabetes island: preliminary impact of a virtual world self-care educational intervention for african americans with type 2 diabetes. JMIR Serious Games 2014;2(2):- [FREE Full text] [doi: 10.2196/games.3260] [Medline: 25584346]

47. Gucciardi E, Fortugno M, Senchuk A, Beanlands H, McCay E, Peel EE. Self-monitoring of blood glucose in Black Caribbean and South Asian Canadians with non-insulin treated Type 2 diabetes mellitus: a qualitative study of patients' perspectives. BMC Endocr Disord 2013;13:46 [FREE Full text] [doi: 10.1186/1472-6823-13-46] [Medline: 24119213]

48. Peel E, Douglas M, Lawton J. Self monitoring of blood glucose in type 2 diabetes: longitudinal qualitative study of patients' perspectives. BMJ 2007 Sep 8;335(7618):493 [FREE Full text] [doi: 10.1136/bmj.39302.444572.DE] [Medline: 17761996]

49. Peel E, Parry O, Douglas M, Lawton J. Blood glucose self-monitoring in non-insulin-treated type 2 diabetes: a qualitative study of patients' perspectives. Br J Gen Pract 2004 Mar;54(500):183-188 [FREE Full text] [Medline: 15006123]

50. Leventhal H, Leventhal EA, Breland JY. Cognitive science speaks to the "common-sense" of chronic illness management. Ann Behav Med 2011 Apr;41(2):152-163. [doi: 10.1007/s12160-010-9246-9] [Medline: 21136224]

51. Meyer D, Leventhal H, Gutmann M. Common-sense models of illness: the example of hypertension. Health Psychol 1985;4(2):115-135. [Medline: 4018002]

52. Torbjørnsen A, Jenum A, Småstuen MC, Arsand E, Holmen H, Wahl A, et al. A Low-Intensity Mobile Health Intervention With and Without Health Counseling for Persons With Type 2 Diabetes, Part 1: Baseline and Short-Term Results From a Randomized Controlled Trial in the Norwegian Part of RENEWING HEALTH. JMIR Mhealth Uhealth 2014;2(4):e52 [FREE Full text] [doi: 10.2196/mhealth.3535] [Medline: 25499592]

53. Ward BW, Schiller JS. Prevalence of multiple chronic conditions among US adults: estimates from the National Health Interview Survey, 2010. Prev Chronic Dis 2013;10:E65 [FREE Full text] [doi: 10.5888/pcd10.120203] [Medline: 23618545]

54. Rollnick S, Miller W. What is motivational interviewing? Behavioral and Cognitive Psychotherapy 1995;23(4):325-334.

55. Roberto CA, Kawachi I. Use of psychology and behavioral economics to promote healthy eating. Am J Prev Med 2014 Dec;47(6):832-837. [doi: 10.1016/j.amepre.2014.08.002] [Medline: 25441239]

Abbreviations

HIT: health information technology
HIV: human immunodeficiency virus
MCC: multiple chronic disease

Original Paper

More Than Telemonitoring: Health Provider Use and Nonuse of Life-Log Data in Irritable Bowel Syndrome and Weight Management

Chia-Fang Chung[1], MBA; Jonathan Cook[2], MDes; Elizabeth Bales[3], PhD; Jasmine Zia[4], MD; Sean A Munson[1], PhD

[1]Department of Human Centered Design & Engineering, Dub Group, University of Washington, Seattle, WA, United States

[2]Division of Design, Dub Group, University of Washington, Seattle, WA, United States

[3]Department of Computer Science & Engineering, Dub Group, University of Washington, Seattle, WA, United States

[4]Department of Medicine, Division of Gastroenterology, University of Washington, Seattle, WA, United States

Corresponding Author:
Chia-Fang Chung, MBA
Department of Human Centered Design & Engineering
Dub Group
University of Washington
423D Sieg Hall, Box 352315
University of Washington
Seattle, WA, 98195
United States
Phone: 1 2066961912
Fax: 1 2065438858
Email: cfchung@uw.edu

Abstract

Background: The quantified self, self-monitoring or life-logging movement is a trend to incorporate technology into data acquisition on aspects of a person's daily life in terms of inputs (eg food consumed), states (eg mood), and performance (mental and physical). Consumer self-monitoring mobile phone apps have been widely studied and used to promote healthy behavior changes. Data collected through life-logging apps also have the potential to support clinical care.

Objective: We sought to develop an in-depth understanding of providers' facilitators and barriers to successfully integrating life-log data into their practices and creating better experiences. We specifically investigated three research questions: How do providers currently use patient-collected life-log data in clinical practice? What are provider concerns and needs with respect to this data? What are the constraints for providers to integrate this type of data into their workflows?

Methods: We interviewed 21 health care providers—physicians, dietitians, a nurse practitioner, and a behavioral psychologist—who work with obese and irritable bowel syndrome patients. We transcribed and analyzed interviews according to thematic analysis and an affinity diagramming process.

Results: Providers reported using self-monitoring data to enhance provider-patient communication, develop personalized treatment plans, and to motivate and educate patients, in addition to using them as diagnostic and adherence tools. However, limitations associated with current systems and workflows create barriers to regular and effective review of this data. These barriers include a lack of time to review detailed records, questions about providers' expertise to review it, and skepticism about additional benefits offered by reviewing data. Current self-monitoring tools also often lack flexibility, standardized formats, and mechanisms to share data with providers.

Conclusions: Variations in provider needs affect tracking and reviewing needs. Systems to support diagnosis might require better reliability and resolution, while systems to support interaction should support collaborative reflection and communication. Automatic synthesis of data logs could help providers focus on educational goals while communication of contextual information might help providers better understand patient values. We also discuss how current mobile apps and provider systems do, and do not, support these goals, and future design opportunities to realize the potential benefits of using life-logging tools in clinical care.

XSL•FO
RenderX

Introduction

Overview

People increasingly turn to mobile phone apps and wearable devices for health management. Nearly 70% of US adults track at least one health indicator for themselves or for a loved one, especially if they suffer from a chronic medical condition; 21% use some form of technology to do so [1]. This year, an estimated 500 million people worldwide will use a health care app [2]. There are already over 5000 self-monitoring apps in Apple's iOS app store alone, including physical activity (eg, Moves [3]), sleep (eg, Sleep Cycle [4]), and food journaling (eg, MyFitnessPal [5]) apps.

Many companies promise to take advantage of cheaper, low-power sensors and ubiquitous connectivity to make patient data about everyday health factors, behaviors, and outcomes available to their health providers. This, they argue, will lead to higher quality and less expensive health care. Partnerships between consumer device manufacturers and health care systems, such as Apple's HealthKit [6] in which Apple collaborated with hospital systems and electronic health record (EHR) vendors, also promise to make the vast array of data collected from these apps available to health providers. To achieve this potential, designers and administrators must consider provider goals, perceptions, and workflows.

Health Information Technology and Life Logs

Self-monitoring—using data manually entered by individuals or objectively collected with sensors in phones or other devices—has proven valuable for patients managing chronic diseases that require behavior change [7,8]. Prior work has identified barriers to using nonelectronic tracking tools, such as paper-based diaries, for managing chronic diseases, either individually or in consultation with providers [9,10]. Electronic diaries have resulted in improved compliance rates, more complete and higher quality entries, and greater user satisfaction in comparison to paper diaries [9,11,12]. Wearable devices and mobile phone apps that enable objective monitoring, such as wirelessly connected pedometers, can further reduce user burdens and improve the completeness and precision of data.

Consumer-focused life logs are a commonly used type of self-monitoring app; these may contain health and wellness data, such as food, physical activity, and mood, as well as nonhealth data, such as location and calendar information. Early life-logging prototypes were envisioned as systems that would capture every aspect of someone's everyday life [13]. The consumer life-logging and personal informatics apps [14] on today's market typically track one or two types of data (eg, the mobile phone app, Moves, records locations, how people travel between those locations, and physical activity) or are designed to help people inspect and manage a particular health concern or other aspect of their life. Mobile apps that track physical activity and increase user awareness of activity levels can increase motivation for behavior change [15,16]. Other systems

have been designed to help end users identify factors that influence their wellness behaviors and outcomes, for example, as discussed in Bentley et al [17] and Kay et al [18]. These life-logging and tracking tools have the potential to make continuous, objective, and precise data available for clinical care. Integration of data that many people already collect using these apps may then enrich medical care without adding new burdens for patients.

While there is increasing interest in incorporating these data into electronic medical records [19], how to integrate these data into provider workflows without increasing provider burdens remains underexplored.

Provider-Patient Communication

Technology tools can also enhance patient-provider communication. Better communication in medical care correlates with better patient adherence [20] and other intermediate outcomes associated with improved health [21]. Many studies have identified important objectives of provider-patient communication, such as creating a good interpersonal relationship, exchanging information, and performing shared decision making [22]. Technology tools can support these objectives. Studies of patient-care management tools for cancer [23] and diabetes [24,25] have shown how technology can support communication around care. In these studies, symptom-tracking data helped patients promote conversations with clinicians and empowered patients to control their ways of interacting with doctors.

The use of technology during clinical visits can also harm provider-patient communication. While paper-based medical records can provide flexibility and a focus point during clinical conversations, the use of computer-based electronic medical records can create barriers to eye contact and communication [26-28].

In recent years, telemonitoring has shown promise for various health management programs, such as reducing readmission rates for heart failure patients [29,30] and for diabetes management [31,32]. However, most telemonitoring systems have been designed with providers as the primary users and designers have overlooked patient roles in providing context and interpreting symptom data [33].

Self-trackers, on the other hand, predominantly use one or more of a variety of consumer-focused life-logging apps and devices available in the marketplace; these have been designed with the self-tracker as the primary user, independent of medical professionals or other experts. Despite these tools not being designed for the patient-provider interaction, around one-third of current self-trackers have shared data from these apps with their health providers [34]. When shared, however, providers rarely engage with this data, which frustrates many of those self-trackers [34]. These aspirations and frustrations are likely to increase as mobile phones gain more sensors, and more people adopt wearable devices and health tracking apps [1,34]. Integrating data from these apps into clinical care may greatly

expand the use of self-monitoring data by medical teams, but at the expense of disrupting current practices and routines.

Research Goals

Overview

In our study, we sought to build upon previous findings to identify specific opportunities and barriers for the use of personal informatics tools in clinical care. We focused on two chronic conditions that are affected by everyday choices: weight management and irritable bowel syndrome (IBS). Over time, both conditions result in high direct and indirect health care costs.

Overweight and Obesity

The prevalence of overweight and obesity are increasing in the United States [35] and worldwide [36]. Overweight and obesity are associated with increased risk of cardiovascular diseases, diabetes, hypertension, certain cancers, respiratory problems, and osteoarthritis [37].

Addressing barriers to change, self-monitoring and strategizing how to maintain lifestyle changes have proven to be effective techniques for weight management. Counseling about changes in diet and physical activity is desirable to patients [38], can promote increases in physical activity [39], and results in significant, sustained weight loss [8]. Despite ample evidence supporting behavior change programs in treating obesity, primary care providers report that inadequate training and lack of time are significant barriers to providing counseling for weight-loss patients [10].

To address the resource intensity of in-person behavioral weight-loss programs, technology-enabled approaches are increasingly common. Mobile phone apps that assist with goal setting and self-monitoring may help overweight and obese patients lose weight [12]. Integration of this consumer-collected data with health care providers' routines may increase and reinforce the efficacy of behavior change efforts, but design and treatment practices to support use of this data are currently unknown.

Irritable Bowel Syndrome

IBS is a chronic functional disorder characterized by the presence of episodic abdominal pain associated with diarrhea and/or constipation. It affects up to 20% of the US population [40]. Each individual patient with IBS has a different set of potential behavioral factors that can trigger a symptom flare-up. Providers currently attempt to identify these individualized IBS symptom triggers by manually scanning paper diaries for food, sleep, activity, and symptom correlations. However, these diaries are typically handwritten with incomplete, disorganized, and unreliable data. Therefore, providers often do not have the time or ability to interpret data from patient diaries [11]. As a result, 62.5% of IBS patients are dissatisfied with the feedback they receive from providers regarding their diaries [41].

IBS patients and their providers need a more efficient and effective way to use the data in these diaries to identify individualized lifestyle modifications that result in bowel symptom reduction and improved quality of life. Limited research has been conducted on the perspectives of IBS providers on the goals for collecting, sharing, and representing life-log data.

Through interviews with 21 health care providers, including physicians, nurses, and dietitians, we contribute an understanding of providers' current practices and constraints for reviewing life-log data and their concerns and needs with respect to this data. Most providers already have, and believe they should have, a role in reviewing patients' life-log data; however, their opinions differ on specific roles and how much this review can support diagnosis, development of treatment plans, and the patient-provider relationship. Providers encounter this data when patients bring in their own journals, unsolicited, or after directing patients to track as part of the diagnosis and treatment process.

While many health providers recognize that patients prefer to use, or already use, consumer-focused life-log apps, these apps offer insufficient flexibility to meet provider needs and do not support collaborative use during patient visits. Consequently, providers rarely review data from patient apps, frequently delegate the review of data to others, and usually ask that patients use provider-preferred tools even if they are less convenient for patients. Finally, though it was not the focus of this study, we identified organizational and policy barriers to provider-preferred workflows for reviewing patient-collected data.

Methods

Overview

To answer these questions, we conducted interviews with medical providers. Our interviews covered their current use of patient-collected data, how they aspire to use this data (if they do), and what facilitates and inhibits such uses.

Participants

We recruited 21 primary care providers through word-of-mouth recruitment with colleagues. We focused on health providers in a large, university-affiliated health system (1). To gather perspectives from providers in other health systems, we also interviewed providers in a second university-affiliated health system (2), a health maintenance organization (1), and one independent dietitian. Many providers also had experience working in other university-affiliated health systems (participant IDs: D02, GM09, GM10, D13, D15, GM21, D17, FM18), another health maintenance organization (2) (participant ID: FM16), and other private practices (participant IDs: D02, FM05, FM06, D11, D12). Some of these organizations were in other states (participant IDs: D02, GM09, GM10, D15, FM16, FM18, D20). Participants included 6 family medicine physicians, 1 behavioral psychologist, 1 nurse practitioner, 5 gastroenterologists, and 7 dietitians (see Table 1). We compensated each participant with a US $30 gift card.

Table 1. Study participants.

Affiliation	Specialty	Participant IDs
University-affiliated health system (1)	Family medicine physician	FM01, FM05, FM06, FM07
	Dietitian	D02, D11, D12, D13, D15
	Gastroenterologist	GM09, GM10, GM14, GM19
	Behavioral psychologist	BP03
	Nurse practitioner	NP04
	Health navigator	HN08
University-affiliated health system (2)	Dietitian	D20
Health maintenance organization (1)	Family medicine physician	FM16, FM18
	Gastroenterologist	GM21
Independent	Dietitian	D17

The family medicine physicians, dietitians, nurse, and behavioral psychologist we interviewed work with patients on a variety of concerns, including IBS and obesity/overweight, while the gastroenterologists work specifically with patients with digestive problems, such as IBS.

Our results describe practices, goals, and barriers experienced by a variety of health providers. We believe the results describe most US health systems, though we note where we identified differences between health systems. Further, because the providers we interviewed practice at a variety of clinic sites, we were able to learn about experiences providing care to patients with diverse backgrounds and socioeconomic statuses. This is important, as personal informatics tools are commonly critiqued, for the most part, as tools for technically savvy, well-off individuals.

Interviews

We conducted an hour-long (range 50-70 minutes) semistructured interview with each participant. We interviewed 10 participants in person and 11 by phone. The interview consisted of three segments intended to help us learn about provider experiences, goals, and concerns about using patient-collected life-log data during patient visits (see Multimedia Appendix 1). In the first portion, we asked whether providers currently review any patient-collected life-log data as part of patient visits. If they did, we probed for the type of data, the clinical conditions for which they use this data, the review process, and their best and worst experiences with the review process. If they did not review patient-collected life-log data, we explored why they did not use this data. Next, we followed up with questions exploring how patient-collected life-log data does and does not currently fit into provider workflows. These questions included inquiries into the benefits that the data and its review offers to the provider and their patients, challenges in reviewing data, provider goals in reviewing the data, and roles in the collection and review process.

To help providers react to specific examples of different types of data, including providers who were less familiar with personal informatics tools, we used three paper prototypes in the interviews. These included a dashboard for a physical activity tracking device currently on the market (Fitbit [42]), a mobile app to help IBS patients track symptoms and triggers (Gut Guru [43]), and Health Report [44], a conceptual app that allows patients to track symptoms between visits and then summarize their data before a clinic visit (see Figure 1, a and b). For remote interviews, we presented prototypes using video chat features or sent screenshots by email.

Figure 1. Patient-collected life-log apps (a, b) and a paper-based journal (c).
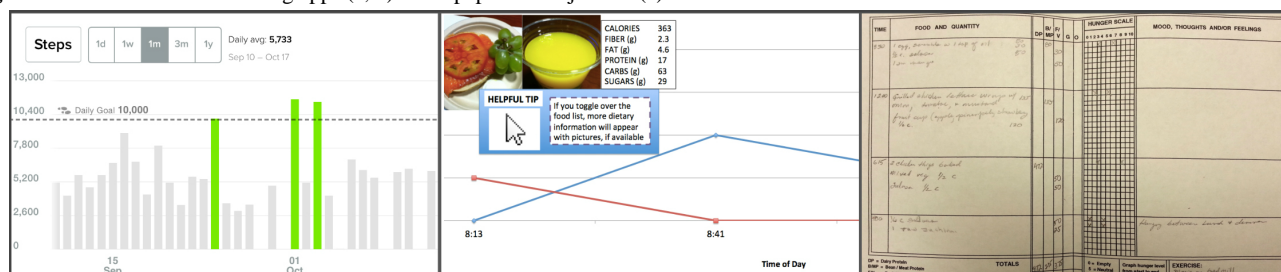
Analysis Techniques

We audiotaped and transcribed all interviews. The full research team conducted an affinity diagram analysis [45] to identify key themes. We transformed the interview transcripts into approximately 700 affinity notes. After several passes inductively organizing these notes into categories, we identified several themes regarding provider-perceived benefits and barriers to use of consumer-oriented, self-monitoring data (see Multimedia Appendix 2).

In addition to our affinity analysis, we coded each transcript through a mix of deductive (based on our research questions and themes identified in prior work) and inductive coding to

capture other emergent themes. Two researchers independently coded the same transcripts and met to resolve ambiguities in the codebook and to add and refine codes reflecting emergent themes. After coding the remaining transcripts, we rereviewed all transcripts to reflect the final codebook. In coding transcripts, we coded especially for goals and barriers to life-log data use.

Results

Overview

Overall, providers saw potential benefits for using life-log data, but rarely engaged with data from the apps many patients already use. Of the various types of everyday life-log data, providers were most familiar with symptom and performance diaries, such as an IBS symptom diary or a weight record. Most providers (except participants N04, HN08) were familiar with various forms of food journals. Some (all dietitians and participants BP03, FM01, FM05, FM16, FM17, GM18) had experiences with encouraging patients to track their physical

activities or with patients bringing in their physical activity records. Providers also mentioned using or encountering other types of life logs, such as sleep logs (participants FM01, BP03, D12), stress logs (participants FM01, BP03, GM09, D15, FM16, GM18), or mood diaries (participants BP03, D15, D17). MyFitnessPal, Weight Watchers [46], and Fitbit were the most common apps that patients had asked providers to look at and review. Table 2 shows a summary of provider attitudes about the benefits and barriers of self-tracking and reviewing tracked data.

Providers noted that even patients tracking on their own—without review by providers—can be beneficial, but that their instruction and review can help patients overcome many obstacles to effective self-tracking. Providers also reported ways that their review of patient-collected data can help achieve their treatment and communication goals. Despite the many benefits that providers have experienced and perceived, barriers to integrating this data into their workflow remain, and these barriers deter them from fully adopting the practice.

Table 2. Summary of results showing provider attitudes about benefits and barriers of self-tracking and reviewing tracked data.

Tracking by patient or provider	Benefits	Barriers and concerns
Patients tracking without provider involvement	Patients can learn about their behaviors, symptoms, progress, and health outcomes. Patients can identify trends and correlations from their data. Patients can become more independent in managing their conditions.	Self-tracking requires extensive patient time and commitment. Patients may not understand what to track and how to track well. Patients may have unrealistic expectations and lose motivation when they cannot immediately reap the benefits of tracking.
Tracking directed by medical team	Tracking overcomes some patient motivational barriers. Tracking provides opportunities for patient education.	Asking patients to track without having providers review the data can send mixed messages. Some providers doubt their ability to advise on tracking; many providers doubt patient ability to review tracking data. Providers may be unfamiliar with currently available tracking tools.
Tracking reviewed by health care providers	Tracking supports diagnosis. Tracking helps personalize treatment(s). Tracking increases patient motivation and accountability. Tracking supports the patient-provider relationship.	Providers have constrained time. Providers question additional health benefits from provider review of tracked data. There is a lack of tracking tool flexibility. There is a lack of tracking tool standardization. There are no established mechanisms for patients to share tracking data with providers.

Benefits of Tracking, Even Without Provider Involvement

Consistent with prior literature on the value of behavioral self-monitoring [8,39], many physicians considered tracking as an opportunity for patients to learn about their behaviors, symptoms, weight-management progress, or other health outcomes, and generally as a way to manage their health. If patients can track, review, and analyze what they do on a daily basis, it "will change how [they] do it" (participant GM10). Participant D15 described how tracking data becomes an "eye opener" for some of her patients. Tracking helps weight-management patients see if, and how, they need to change their food intake and can help IBS patients identify how certain symptoms correlate with specific foods.

Some physicians noted that as people develop good tracking practices, they become capable of identifying trends and correlations between their behaviors and symptoms, and become more "independent of physicians" (participant GM09). When people in a weight-loss program are already tracking their diet and activity, participant D12 worries less about their chances for success: "...they are a little bit more motivated, they are going to become more educated in that, and that is going to help them meet those goals."

Provider-Perceived Patient Barriers to Tracking

Providers also described many barriers that prevent patients from effective tracking. Some tracking requires a regular commitment of time and effort, especially for data that cannot currently be logged automatically and unobtrusively, such as food intake. Therefore, many patients do not adhere to tracking

with sufficient frequency or detail to make this data clinically useful. Patients might also not understand what to track and how to track well. Some patients may track the wrong data or may consistently not track well or accurately enough to gain benefits; these patients may lose their motivation for continued tracking.

Patients also sometimes have unrealistic expectations for how quickly they will receive benefits from tracking. They become discouraged when they do not achieve them. Even when patients track regularly and track the right factors, the apps they use do not always help them draw actionable conclusions from this data. For example, some IBS patients may expect to identify specific foods that are symptom triggers, when specific ingredients may be the trigger.

To address these barriers, patients may require alternative tools or guidance from health experts on deciding what to track, as well as how to interpret and act upon the data.

Benefits of Medical Team-Directed Tracking

Providers find that simply encouraging patients to track the data, and later asking them how it is going, overcomes some motivational barriers to tracking and managing chronic illness. As noted in the previous section, additional coaching on what to track, how to track, and how to review the data can help patients learn to track the right data in the right ways. This can also positively impact patients' confidence in their ability to benefit from tracking. Many providers—physicians, nurses, and dietitians—considered themselves responsible for educating patients on how to track and make use of data:

> We play a role of educating patients about the use of the data, so that they don't overreact or underreact. Putting them into our perspective, and then being able to use that across the board. [Participant N04, who uses symptom diaries]

Benefits of Reviewing Tracked Data

Overview

Provider review of patient-collected data can offer a variety of further benefits. Many providers reported using patient-collected life-log data to make, modify, or confirm diagnoses. Others use the data to tailor treatment plans to patients' routines.

Supporting Diagnosis

For providers working with patients in weight-management or chronic disease-management programs, diagnosis and treatment is usually a multi-step and multi-provider process. They rely on patient-collected data across multiple visits to make and adjust their diagnosis and treatment plan. In most weight-management programs, primary care physicians often diagnose overweight or obesity by assessing body mass index (BMI), waist measurement, and other health risk indicators. After diagnosis, patients are often referred to a dietitian for diet and exercise management. Participant D15 described her process during weight-management dietary consultations:

> At an initial visit I probably spend 15-20 minutes on reviewing what they eat [using a diet recall] and pen that down. When they bring a food record in we'll

> spend about the same amount of time going over that. Sometimes people will send me records then I follow up by emailing them and asking questions. Usually what I'll do is look at the whole thing to pick typical days and a few other random days so that I can tell the difference between calories. [Participant D15]

Participant GM09 described a similar approach for using food and symptom journals with potential IBS patients:

> If I see a patient for an initial encounter and if I think that they have IBS, I would give them a task to collect some data then see them back in a couple of months. Then I'll review it with them and see if together we can come up with some trends or interventions that might be beneficial for them. Once we have a diagnosis and institute a treatment plan then my goal in each visit is to see if that's working or if we need to adjust our management plan. [Participant GM09]

Participant D2 described an experience working with a patient who had trouble determining the reason for her weight increase. After she started tracking her food intake and exercise, participant D2 and the patient reviewed the food log. Together, they found out how many calories came from the bottle of wine the patient drank each day. After identifying and acting on this opportunity for change, the patient began to lose weight.

With multidisciplinary teams becoming increasingly important in primary care, providers also use patient-collected data to support decisions about when to involve other medical team members. Participant FM01 imagined that noticing poor sleep hygiene from a patient's sleep diary might lead him to refer the patient to a behavioral psychologist to evaluate any psychological etiology for insomnia. Participant D12 also believed that access to patient-collected data by all medical providers allowed for a more cohesive team approach for patient care:

> [If patient-collected data indicates that other psychological factors, not just dietary intake, are affecting symptoms] the dietitian gets support too. Because you got other providers working with the patient too. [Participant D12]

Personalizing Treatment

Reviewing self-monitoring data together also provides opportunities for providers to learn about their patients. Understanding patient preferences and routines helps providers shape care to individual needs. Participant D13 talked about her experience working with people who described dieting their whole lives, but who were still unable to lose weight. By working with them to understand their routines, she was able to tailor the diet suggestions based on their individual diet constraints or cooking habits.

IBS diagnosis and treatment is a complex process that requires excluding other conditions that produce similar symptoms, categorizing patient symptoms, and determining individualized, heterogeneous triggers. It also requires identifying which symptoms a patient most wants to address and what changes he or she is, and is not, willing to make to manage them. To help with this process, health providers want tracking tools that

are more flexible and that support tracking the symptoms most important to their patients:

> It would be helpful if the patient has another symptom that they think might be related and be able to track that symptom as well as [what they are already tracking]. [Participant GM10]

Communicating around shared symptom and behavior logs can also help providers learn about their patients' priorities, both for symptoms and lifestyle.

Increasing Motivation and Accountability

Directing patients to track data and coaching them on what to track benefits many patients, but many others continue to struggle with motivation. Providers find that reviewing data with these patients can show them "why it's valuable for [them] to collect those data" (participant D11), leading to increased patient engagement with the tracking activity, as well as with the overall treatment plan. This is especially the case when patients are unable to draw meaningful conclusions from the data on their own. Going to the effort of collecting data but not having it lead to improved health or a reduction in symptoms is a frustrating, tiring process. Participant D02 emphasized the importance of reviewing data with patients in this situation:

> It's more helpful if you have someone to review it with, because otherwise it might look like "Why am I doing this to myself? I'm just taking all this time for nothing." [Participant D02]

Consistent with Mohr et al [47], providers also felt their review of patient data could increase patient accountability and adherence to the treatment plan. One of participant D12's patients noted, "I know I have to turn it in so I'm going to eat healthier."

Supporting the Patient-Provider Relationship

Overview

Reviewing self-monitoring data also helps providers communicate and build relationships with their patients. Effective provider-patient communication supports information exchange and shared decision making, which may increase patient knowledge about their health status and adherence to the treatment [22].

Learning About Patients

Many providers use patient-collected data and the review process to learn about their patients, to "get an idea of what's going on in their life" (participant GM10). While this information can help identify alternative sources or triggers of certain symptoms, it also can reveal unarticulated patient values and goals. Data common in consumer life-log apps, but not in many provider-preferred tools—"What do you do? Do you live by yourself? How long is your commute? Who else lives with you? What kinds of obligations to your time do you have?" (participant D13)—reveal constraints and opportunities for change. Participant FM06 thought she could discover "what the patient cares about" from tracked data and conversations about that data, and then use this information to motivate the patient to stay in the program. For example, participant D20 asked a

patient to record the context of eating and found he ate more when he had peer pressure from cousins and friends. Therefore, instead of just telling him to eat less, they brainstormed ways to improve his diet without sacrificing his social life.

Facilitating Discussion and Managing Visits

Many providers use patient-collected data to facilitate discussions during visits. If providers have access to the data prior to or at the beginning of the visit, they can plan the visit agenda around patient concerns or have a topic to initiate the conversation. For participant D02, the tracked data is particularly useful when patients do not have, or have trouble articulating, a clear reason for a visit:

> If they don't know what the problem is, I at least have something to look at, and I can identify where to start to ask questions around rather than having a million things I can ask about but not knowing if any is relevant. [Participant D02]

Some providers prefer to have the data ahead of time so they can better prepare for the visit, particularly when a patient collects a considerable amount of data between visits:

> If I had seen this [report] beforehand, this would be really nice for me to know what she is planning on coming in and what she wants to talk about so it doesn't catch me by surprise, so I can prepare for it too. [Participant D12]

Others felt they would rarely have time to review patient data before a visit, so they preferred to engage with this patient data only during visits.

Some providers use patient journals to facilitate and create a record of conversations during visits. Participant D13 showed us her favorite paper-based food journal (see Figure 1, c) and explained how she uses it to facilitate her conversations with patients. During a visit, she highlights certain columns to emphasize main points of the conversation or for follow-up. She crosses out other columns to alleviate unnecessary patient concerns and to help focus the conversation. Patients can then take the annotated record and use it to reflect on their behavior. This gives patients an artifact that supports their memory of the conversation and can help them journal more efficiently in the future.

Participant BP03 has his patients practice cognitive behavioral therapy at home and send him their thought records before visits. When they are together in the clinic, participant BP03 shares his computer screen with his patients and they review the data together. This helps participant BP03 and his patients better understand each other's focus and correct any misunderstandings right away.

Barriers to Using Tracked Data in Clinical Care

Overview

Despite the benefits of reviewing patient tracking plans and patient-collected data in clinical care, providers encounter many challenges when they try to use this data in their current practice. Primary care and gastroenterology physicians found it challenging, if not impossible, to review large amounts of data

during short visits and they lack incentives to review it outside of office visits. They also questioned whether they have the appropriate expertise to review the data. Therefore, they prefer to delegate reviewing the data to other medical team members, such as dietitians. These providers typically have longer visits with patients and more expertise and experience in identifying problems using life logs, especially food diaries. However, dietitians experience their own time and workflow constraints to reviewing the data, complicated by electronic tools that do not support their needs. In the next sections, we review these barriers and concerns in detail.

Lack of Time

Across providers we interviewed, the common clinic appointment in family medicine or gastroenterology lasts 15 to 20 minutes, leaving less than 5 minutes—more often 1 to 2 minutes—for a provider to review self-tracked data. Many are skeptical about what they can meaningfully achieve in that time. For example, participant FM07 said she does not have enough time during a visit to explain what the data means, and so she chooses not to review it at all.

Some physicians we interviewed believe reviewing data between visits is valuable, but they also feel this work is not recognized and thus they cannot allocate much time to it. Participant FM05 reported he could only spend 5 to 10 minutes per day reviewing patient-collected data across all patients (around 20 patients per day). Current workflows and incentive structures pose "a time barrier that discourages me from reviewing" patient-collected data (participant GM09).

Compared to physician visits, the typical patient visit with a dietitian lasts 30 to 60 minutes. This causes many physicians to delegate the review of self-monitoring data to dietitians and focus on other topics during their short visits. Dietitians normally spend 15 to 20 minutes reviewing tracked data—predominantly food journals, sometimes along with a physical activity or symptom diary—with patients and consider it a valuable part of their consultation.

Many dietitians also work with patients on their tracked data outside of clinic visits. Patient portals and other tools for online communication enable dietitians to review data, discuss barriers to tracking, and recommend changes to a treatment plan on an ongoing basis. Many dietitians believe this is the most effective way to help people to manage their diet for either weight management or IBS symptom control; however, they are hesitant to encourage this practice because this work is unbillable and unpaid across hospital systems. Providers are not normally paid for phone calls, emails, or any electronic communication outside of patient visits. Participant BP03 said the incentive structure "has a perverse, mixed message: collect the data but you don't have time to do it."

Questions About Expertise and Benefits Offered

Most physicians reported that they are often the first person to see patient-collected data, even if they have doubts about their time and expertise to engage with it. Participant GM09 notes that there is often no alternative: "There is no one (to help me review the data). I review the data myself."

Facing time constraints and a lack of expertise, many providers prefer to refer patients to dietitians, when possible. Dietitians have more expertise reviewing food journals and the review activity is "more in line with the normal interaction" (participant GM09). Participant GM14 shared that even though she has a degree in nutrition she does not consider it her strength to understand food types and nutrients efficiently:

> They (dietitians) ask all of those questions and I don't. I'm not good at that. I could probably muddle my way through it but it wouldn't be efficient and I wouldn't get as good information out of it. [Participant GM14]

This support, though, is not always available. While participant FM18 used to refer patients to dietitians for food-related concerns, there are currently no dedicated dietitians in the health maintenance organization (1), so he conducts this review himself.

Other physicians, however, find it more effective to review the records themselves rather than to delegate to other providers. Now that he reviews the data himself, participant FM18 feels that, as the only member of the medical team with regular and consistent interaction with patients, he is in the best position to use the data to offer advice and to enhance his relationship with patients. Participant GM19 also used to refer patients to dietitians, but patients rarely followed up with them. Now he reviews patient-collected data and offers feedback himself.

Lack of Flexibility

Tailoring tracking activity and a treatment plan for an individual patient is important for both weight management and IBS. However, providers reported that current consumer apps and provider tools, including mobile phone and Web apps, do not provide sufficient flexibility to address the needs of both patients and providers. For example, for some patients it is necessary to track a myriad of factors and symptoms in detail, but others might only need to focus on a certain type of food. According to participant D13, systems that require or encourage patients to track and review more data than necessary often discourage patients from tracking. Participant FM01 said a system should "have the ability for the individual physicians to tailor it to the way they practice," otherwise it is difficult for providers to integrate it into their clinical workflow.

Lack of Standardization

One way patients achieve some flexibility is by choosing among the many apps available in the marketplace. Paradoxically, this creates problems of data standardization when they try to share this data with their health providers. For some health data, such as glucose level, blood pressure, and symptom history, providers use standardized forms or applications. This standard, consistent presentation facilitates their accurate review of data in a short amount of time. However, consumer-oriented life logs often lack a standardized format, or when they do standardize the data, they reduce it to a factor neither the patient nor the health provider cares about (eg, calories for IBS patients). Participant FM16 wanted to use "physical activity as vital signs" in her practice, but found it difficult to compare or to define "activity level" among various types and levels of physical activities.

Some standardized forms for food journaling have been developed for provider use and have been adopted in practice. However, these forms are often difficult for patients to use (cf, Cordeiro et al [48] and Tsai et al [49]) in social settings and to carry around, which leads to incomplete records and recall bias. Participant D15 describes her frustrating experiences with standardized food journals:

> *We had food records as little booklets...but the problem is, how do we get that back to us? They either had to mail it to us, or they had to come in and drop it off, or, if we were seeing them again they could bring it. Half the time people forget to bring it. So, it became easier to just say, "Okay, why don't you record it in any format you want and then either send it to me or bring it to me."* [Participant D15]

Like participant D15, many providers do not provide patients with standardized forms unless patients request them. However, when patients track using their own paper or electronic diaries, this creates challenges for providers to review this free-form, inconsistently formatted data. Free-form data and consumer app that limits on data export also prevent providers from generating and viewing meaningful, actionable summaries.

While needs vary according to provider goals, providers all need different, typically more summarized, views of the data than do their patients. For example, while many patients ask providers to review exercise logs, most providers have difficulty efficiently interpreting heterogeneous, detailed physical activity logs. They find that having patients verbally summarize the logs is still a more efficient way to gain insights from it.

Lack of Mechanisms for Sharing With Providers

Many life-log apps do not offer data export features, other than application programming interface (API) access or sharing to social network sites. This creates barriers to integrating this data into provider practices, especially when they want to be able to review it on a day-to-day basis or during patient phone calls. Participant D15 has better experiences sharing data from paper-based diaries, which can be photographed or scanned, than with current mobile phone apps:

> *[If it's recorded on paper] I can keep it and look at it. If they email it to me I can go back and reference it. If it's on somebody's phone app then I can't.* [Participant D15]

While some providers prefer to conduct in-depth reviews of data only during clinic visits, they also report that the mere potential to access the data remotely would help them achieve more benefits from life-log data, such as increased accountability:

> *[Patients] will not be held accountable for it because there would be no way I could see it before [the] appointment other than [if] they were emailing it in.* [Participant D12]

As a result, some providers hesitate to encourage patients to use life-log apps and prefer that patients regularly email records or bring paper forms to visits.

Discussion

Principal Findings

Providers had varying goals and needs for using and reviewing life-log data (see Table 3). Everyday life logs can support diagnosis, prioritization for follow-up, patient education, patient engagement, and treatment. These data, particularly from provider-directed tracking, are typically used to determine symptom triggers, to guide patient self-monitoring and treatment, and to motivate patient adherence to diagnosis and treatment plans. This set of patient and provider goals, and corresponding design needs, for IBS and weight management is largely consistent with those identified in work with other chronic conditions [24,31,50].

Life-log data, however, also offer more insight into patients' lives and priorities, and thus offer new opportunities to support patient-provider communication, better patient-provider relationships, and more patient-centered treatment plans. These data may include nonhealth data, such as location histories and calendar information, that can help providers learn about patient values, constraints, and goals, as well as help facilitate discussions and help patients and providers manage clinic visits. Currently, this use is almost exclusively at the patient's initiative, and is largely unsupported by current app design.

Table 3. Benefits providers achieve with life-log use in clinical care, as reported for irritable bowel syndrome and weight management.

Provider-perceived benefits	Uses of life-log data	Initiators
Diagnosis support	What and when do patients log?	Health providers
	What triggers symptoms?	
	Do patients track the right things in the right ways?	
Treatment design and planning	What are patient routines and priorities?	Patients
	Do patients adhere to treatment plans?	
Provider-patient relationship building	What do patients value in terms of lifestyle choices?	Health providers or patients
	What are important aspects in the context of patient health behaviors?	

Most of these goals are not well supported by currently available commercial apps, and each set of goals drives different design requirements and opportunities. In this section, we discuss how varying provider goals affect tracking activities and needs, and what design opportunities support each.

Provider and Patient Goals Affect Tracking Needs

Overview

Systems supporting the use of patient-collected data in clinical settings should take into account varying provider goals and tracking needs, such as diagnosis, education, motivation, and relationship goals. Standardization of data formats, accompanied by personalization of which data are collected and how they are presented, will help to support these goals while preventing data from overloading providers and patients [51].

The various patient and provider goals (eg, diagnosis, education, agenda setting, and relationship building) require different balances of personalization and standardization.

Supporting Diagnosis Versus Supporting Interaction

When developing potential treatment plans, providers look for detailed information during their review of patient self-monitoring data. If their goal is to identify specific triggers for symptoms, they look for correlations between factors, whereas if providers are monitoring a symptom or outcome, they try to identify trends and outliers in the data. For example, when working with patients in weight-management programs, providers need precise portion, nutrient, and calorie information to be able to make diet recommendations. When working with IBS patients, providers focus more on common food triggers for IBS symptoms: lactose, fiber, and fermentable, oligo-, di-, monosaccharides and polyols (FODMAP) foods.

However, when providers use patient-collected data to support their interactions with patients, they often look to understand patient goals and priorities. For example, if a patient's goal is to walk 10,000 steps per day, the provider may focus on how frequently the patient meets this goal. Then they can start a conversation about the patient's experience, give suggestions, and discuss remaining barriers. What patients emphasize in symptom logs or food diaries can help providers identify patient priorities and routines about which patients are and are not flexible. Therefore, systems to support patient-provider interactions should provide different granularity of summaries of tracking data that helps providers effectively focus on important facts alongside patients' subjective experiences of this data.

Educating Patients Versus Engaging Patients

Providers play a role in educating patients about their symptoms, health status, tracking, and outcome management. A system can support this educational role. Many traditional telemonitoring systems allow providers to leave feedback on patient-collected symptoms outside of visits [24,33] or to facilitate patients in making sense of their data [31]. Consumer life-log tools have rarely been designed to support this collaborative education, and there are exciting opportunities to do so. For example, automatic synthesis of food journal entries may help providers explain what FODMAP foods are to a newly diagnosed IBS patient, using example foods the patient commonly eats. They could also then discuss categories to focus on, and customize the app to provide feedback on this.

For weight management, many systems support automated calculation of calorie intake and expenditures within a day or over time, helping people see the effects of their choices [49]. These systems could better support patient-provider conversations around diet plans by allowing providers to demonstrate how small differences in choices in the prior weeks could have led to different calorie balances. Such an app could then allow providers to save these simulations as a nutrition plan for the upcoming week.

Providers also use their understanding of patient values, routines, life events, and how tracking fits in with these contextual factors to help keep patients motivated to follow their tracking or treatment plans. For example, participant D13 had an experience when her patient did not fill in the record because her daughter was sick and she did not have time to cook. Seeing this reflected in the record prompted participant D13 to have a supportive conversation to comfort the patient rather than reminding her of the importance of tracking. Therefore, systems to support this provider goal should be capable of collecting and communicating relevant contextual information. While systems to support diagnosis and disease management require high-reliability-tracking technology [52], use of engaging, but perhaps less reliable, consumer-centric tracking apps might offer better support for patient-provider communication around routines, values, and priorities. Prior work (eg, Grönvall and Verdezoto [53]) has noted the importance of routines and values. The ability to use current consumer-centric tools to identify this information, however, has not been previously articulated, at least not in support of some provider goals and for collecting the data; presenting data remains a barrier.

Limitations of Current Systems and Design Opportunities

Overview

Providers who have experience using various tracking tools report that some tools are better at supporting their goals than others. Traditional paper-based diaries still play an important role in many provider-patient interactions, while computer- and mobile phone-based systems provide features to support remote interaction and dynamic representation of data. By understanding the benefits of each system, designers can create life-logging and other self-monitoring apps that will better support provider needs.

Dynamic Representation Versus Flexibility

Computer-based systems have the potential to synthesize data automatically and present dynamic summaries. This helps providers and patients identify trends, correlations, and regular or irregular events during the tracking period. However, most current systems do not give providers the flexibility to adjust tracking plans for what to track and how records and summaries are presented; they do not generate summaries in formats that providers need. As a result, patients often spend considerable effort collecting irrelevant information or collecting potentially valuable information that providers never fully review. On the other hand, paper-based records preserve the flexibility for providers to adjust what data a patient tracks and how they view individual records. Without the resources to enter these records into a standardized database, however, it is impossible to generate summaries automatically or to correlate them with

other data. When patients are diligent in tracking and they bring this data to visits, providers must quickly "eyeball" the data and summarize it themselves.

Future tracking systems should allow providers and patients to customize the data that they track and the ways in which it is presented. Many telemonitoring systems have been designed to allow patients to transmit tracked data to their remote health care team members in a standard format [54]. However, these systems have been designed primarily with the provider's objectives in mind, collecting data and health outcomes at intervals that are meaningful and practical to health care providers. These systems rarely integrate with patient-preferred data collection tools—commercial mobile phone apps or wearable devices—and thus create unnecessary barriers for patients to integrate the tracking process in their everyday routines. Some do not even provide interfaces for patients to view their data. When patients cannot benefit from tracking without involving their health providers, it is difficult for them to stay engaged with the tracking activity [55].

Previous efforts to integrate consumer health data flexibly into health providers' routines have largely failed. Google shut down Health [56] and Microsoft has not reached critical mass with HealthVault [57,58]. Weaknesses included a lack of integration of consumer data into these systems and lack of subsequent use in health care provider workflows [58,59]. Newer frameworks such as Apple HealthKit, Microsoft Health [60], and Google Fit [61] allow interoperability and data portability between different consumer apps and provider tools. They address some of the earlier concerns about poor integration, and may become an important enabling infrastructure by allowing patients and providers to customize tracking tools for an individual based on his/her routines, preferences, and needs while still presenting the data in a consistent, familiar interface for providers.

We caution, however, that while modern health data-sharing systems and frameworks may enable better data sharing for diagnosis and treatment, they are in many regards still less capable than paper-based tracking tools. Their highly standardized formats lose some of the contextual richness that supports patient-provider relationships, communication, and visit management. These formats are also less likely to support the collaborative interactions patients and their providers have around what and how to track using paper journals. For example, while new health frameworks enable consumers to import data from many different tracking tools and share with an app of their provider's choice for review, this unidirectional process reduces opportunities for providers and patients to collaborate on how to use a tool and which specific data to track.

PatientsLikeMe [62] tries to bridge digital and paper formats by allowing patients to track some health factors digitally and then print Doctor Visit Sheets [63,64]. One-third of respondents in a 2010 survey of PatientsLikeMe users reported using Doctor Visit Sheets to support health care visits [64], but none of the providers in our study had encountered one of these summaries. This suggests that the tracking and summarization in PatientsLikeMe works for some users and some conditions, but not those in our study. Tracking in PatientsLikeMe best supports medication and side-effect tracking, with diet tracking limited

to recording vague, nonspecific changes such as "diet modification" for obesity or "low FODMAP diet" for IBS. It also only supports limited integration with other apps or devices, supporting only Fitbit. Similar to modern health data-sharing frameworks, the visit sheets can lose much of the context captured in consumer health and life-logging apps. While this tracking may be useful for understanding the efficacy of medications or broad treatment approaches at the population level, the current capabilities are of limited use to patients seeking to manage IBS, weight, or similar conditions.

Supporting Remote and Face-to-Face Interaction

Providers want self-monitoring tools and their organizations to provide better opportunities for patients and providers to communicate and review data in person and remotely. Patients typically bring the data to their office visits, where there are many challenges to reviewing the data together. In many clinic rooms, arrangement of computer screens does not allow both the provider and patient to review the data. When providers review data on their screen or on a borrowed mobile phone, it creates barriers to eye contact and communication similar to the challenges of using a computer to review electronic medical records (cf, Alsos et al [26] and Chen et al [27]).

Providers often need to discuss, review, and ask questions about life-log data in a collaborative fashion, which is different from when they interact with medical records by themselves and then explain the results to patients. Therefore, instead of designing interfaces and rooms to minimize attention to digital devices [26], designers wishing to support life-log use in clinical settings should provide mechanisms to support shared, collaborative review and interpretation. Sharing life-log app data on a mobile phone, as is often necessary when patients use consumer apps, also means that limited information can be displayed at one time. Paper-based diaries give providers and patients the flexibility to lay out and rearrange different pages during face-to-face interactions, but they limit synthesis or remote sharing and review.

Clinic administrators and app designers might explore screen-casting capabilities. Sharing mechanisms have been described in previous research [24] and by providers that we interviewed (participants D02, D15)—patients either bring their own device or share their passwords with providers. In contrast, screen casting would allow patients and providers to review more data on a shared screen and still allow patients to retain "ownership" and control of their data. Collaboratively annotated individual records or summaries would also provide patients with an artifact of the conversation that the patient could use to support their tracking and treatment between visits [24].

Organizational Support

Reviewing patient-collected life-log data can help providers manage visits and develop or adjust treatment plans. It can also support patient-provider communication. By helping providers learn about patient values, goals, and constraints, and by offering them real examples from patient data to use in conversations, life-log data supports patient-provider communication, which supports better patient outcomes [21]. On the other hand, health care decision makers report barriers to adopting

technology-based tools, such as limited budget and resources or concerns about reimbursement policies [65]. Similar to the providers in our study, primary care teams in other recent studies report lacking the time, education, and compensation necessary for them to review dietary data and provide dietary consultation, creating barriers to adopting electronic-based dietary assessment tools [50]. To achieve the benefits and overcome the barriers associated with provider review of tracking data, policy makers and health system leaders should evaluate whether incentive structures and clinic visit workflows can be revised to enable potential benefits to be achieved more reliably.

Our study provides one of the first in-depth examinations of clinician attitudes toward, and experiences with, using life-log data in clinical care. The perspectives of these front-line providers will help guide the design of life-log tools, support providers and patients with accessing this data, and steer policies regarding integration of both into current processes. Future *in situ* observational studies could provide more-detailed insights into the process of reviewing life-log data and its influence on clinical workflow.

Conclusions

In this study, we characterized the benefits health providers gain through use of life-log data, as well as the barriers they face in achieving these benefits. In particular, we identified two broad categories of life-log data use in IBS and weight management:

supporting diagnosis and treatment, as well as building patient-provider relationships in ways that support open communication and tailoring treatment plans to patient priorities and routines.

Regardless of goals, providers face many barriers to integrating life-log data into current practice. Even though many patients first bring their data to their physicians, physicians prefer to refer these patients to providers with more accommodating workflows and better-matched expertise. Dietitians, on the other hand, have to work with patients outside of clinic time, without extra reimbursement, to make the best use of the data. Current apps hinder effective tracking and data sharing by not offering flexibility to tailor tracking plans, by not supporting standard data types and summaries for provider review, and by not providing a mechanism for provider review on a day-to-day basis.

New health data-sharing frameworks, such as Apple HealthKit, Microsoft Health, and Google Fit, promise to integrate consumer-collected life-log data automatically, including data sensed by the companies' own platforms. Apple has worked with the Mayo Clinic and other leading health care organizations to design for integration and summarization from the start [66], but these improvements may not be enough to ensure successful adoption. Designers of health apps should be aware of the range of benefits life-log data can offer, and of the different fidelity and presentation requirements for each goal.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Semistructured interview protocol.

[PDF File (Adobe PDF File), 38KB - jmir_v17i8e203_app1.pdf]

Multimedia Appendix 2

Affinity diagram labels.

[PDF File (Adobe PDF File), 46KB - jmir_v17i8e203_app2.pdf]

References

1.   Fox S, Duggan M. Pew Research Center. Washington, DC: Pew Internet and American Life Project; 2013 Jan 28. Tracking for health URL: http://www.pewinternet.org/files/old-media/ /Files/Reports/2013/PIP_TrackingforHealth%20with%20appendix.pdf [accessed 2015-01-03] [WebCite Cache ID 6VK9cmf7I]
2.   Patient Apps for Improved Healthcare: From Novelty to Mainstream. Parsippany, NJ: IMS Institute for Healthcare Informatics; 2013 Oct. URL: http://www.imshealth.com/deployedfiles/imshealth/Global/Content/Corporate/IMS%20Health%20Institute/ Reports/Patient_Apps/IIHI_Patient_Apps_Report.pdf [accessed 2015-01-03] [WebCite Cache ID 6at2h0TFs]

XSL•FO

RenderX

3.   Moves. URL: https://www.moves-app.com/ [accessed 2015-08-10] [WebCite Cache ID 6aiDAGMZN]

4.   Sleep Cycle. URL: http://www.sleepcycle.com/ [accessed 2015-08-10] [WebCite Cache ID 6aiDGrL1n]

5.   MyFitnessPal. URL: https://www.myfitnesspal.com/ [accessed 2015-08-10] [WebCite Cache ID 6aiDUaMHX]

6.   HealthKit. URL: http://www.applehealthkit.com/ [accessed 2015-08-10] [WebCite Cache ID 6aiDroN8i]

7.   Heitkemper MM, Jarrett ME, Levy RL, Cain KC, Burr RL, Feld A, et al. Self-management for women with irritable bowel syndrome. Clin Gastroenterol Hepatol 2004 Jul;2(7):585-596. [Medline: 15224283]

8.   US Preventive Services Task Force. Rockville, MD: US Preventive Services Task Force; 2012 Jun. Final Recommendation Statement. Obesity in Adults: Screening and Management URL: http://www.uspreventiveservicestaskforce.org/Page/Document/RecommendationStatementFinal/obesity-in-adults-screening-and-management [accessed 2015-01-03] [WebCite Cache ID 6VKInTHJV]

9.   Jhaveri M, Lee E. Performance of electronic diaries in diabetes clinical trials measured through overall satisfaction of site coordinators. J Diabetes Sci Technol 2007 Jul;1(4):522-530 [FREE Full text] [Medline: 19885115]

10.   Kushner RF. Barriers to providing nutrition counseling by physicians: a survey of primary care practitioners. Prev Med 1995 Nov;24(6):546-552. [doi: 10.1006/pmed.1995.1087] [Medline: 8610076]

11.   Heinonen R, Luoto R, Lindfors P, Nygård CH. Usability and feasibility of mobile phone diaries in an experimental physical exercise study. Telemed J E Health 2012 Mar;18(2):115-119. [doi: 10.1089/tmj.2011.0087] [Medline: 22283356]

12.   King AC, Hekler EB, Grieco LA, Winter SJ, Sheats JL, Buman MP, et al. Harnessing different motivational frames via mobile phones to promote daily physical activity and reduce sedentary behavior in aging adults. PLoS One 2013;8(4):e62613 [FREE Full text] [doi: 10.1371/journal.pone.0062613] [Medline: 23638127]

13.   Hodges S, Williams L, Berry E, Izadi S, Srinivasan J, Butler A, et al. SenseCam: a retrospective memory aid. In: Proceedings of the 8th International Conference on Ubiquitous Computing (UbiComp'06). Heidelberg, Germany: Springer-Verlag Berlin; 2006 Presented at: 8th International Conference on Ubiquitous Computing (UbiComp'06); September 17-21, 2006; Orange County, CA p. 177-193. [doi: 10.1007/11853565_11]

14.   Li I, Dey A, Forlizzi J. Understanding my data, myself: supporting self-reflection with ubicomp technologies. In: Proceedings of the 13th International Conference on Ubiquitous Computing (UbiComp '11). New York, NY: ACM; 2011 Presented at: 13th International Conference on Ubiquitous Computing (UbiComp '11); September 17-21, 2011; Beijing, China p. 405-414. [doi: 10.1145/2030112.2030166]

15.   Consolvo S, Everitt K, Smith I, Landay JA. Design requirements for technologies that encourage physical activity. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '06). 2006 Presented at: SIGCHI Conference on Human Factors in Computing Systems (CHI '06); April 22-27, 2006; Montréal, QC p. 457-466. [doi: 10.1145/1124772.1124840]

16.   Consolvo S, McDonald DW, Toscos T, Chen MY, Froehlich J, Harrison B, et al. Activity sensing in the wild: a field trial of UbiFit Garden. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '08). New York, NY: ACM; 2008 Presented at: SIGCHI Conference on Human Factors in Computing Systems (CHI '08); April 5-10, 2008; Florence, Italy p. 1797-1806. [doi: 10.1145/1357054.1357335]

17.   Bentley F, Tollmar K, Stephenson P, Levy L, Jones B, Robertson S, et al. Health mashups: presenting statistical patterns between well-being data and context in natural language to promote behavior change. ACM Trans Comput Hum Interact 2013 Nov 01;20(5):1-27. [doi: 10.1145/2503823]

18.   Kay M, Choe EK, Shepherd J, Greenstein B, Watson N, Consolvo S, et al. Lullaby: a capture and access system for understanding the sleep environment. In: Proceedings of the 2012 ACM Conference on Ubiquitous Computing (UbiComp '12). New York, NY: ACM; 2012 Presented at: 2012 ACM Conference on Ubiquitous Computing (UbiComp '12); September 5-8, 2012; Pittsburgh, PA p. 226-234. [doi: 10.1145/2370216.2370253]

19.   Fitzpatrick G, Ellingsen G. A review of 25 years of CSCW research in healthcare: contributions, challenges and future agendas. Comput Support Coop Work 2012 Jun 21;22(4-6):609-665. [doi: 10.1007/s10606-012-9168-0]

20.   Zolnierek KB, Dimatteo MR. Physician communication and patient adherence to treatment: a meta-analysis. Med Care 2009 Aug;47(8):826-834 [FREE Full text] [doi: 10.1097/MLR.0b013e31819a5acc] [Medline: 19584762]

21.   Street RL, Makoul G, Arora NK, Epstein RM. How does communication heal? Pathways linking clinician-patient communication to health outcomes. Patient Educ Couns 2009 Mar;74(3):295-301. [doi: 10.1016/j.pec.2008.11.015] [Medline: 19150199]

22.   Ong LM, de Haes JC, Hoos AM, Lammes FB. Doctor-patient communication: a review of the literature. Soc Sci Med 1995 Apr;40(7):903-918. [Medline: 7792630]

23.   Patel RA, Klasnja P, Hartzler A, Unruh KT, Pratt W. Probing the benefits of real-time tracking during cancer care. AMIA Annu Symp Proc 2012;2012:1340-1349 [FREE Full text] [Medline: 23304413]

24.   Aarhus R, Ballegaard SA, Hansen TR. The eDiary: bridging home and hospital through healthcare technology. In: Proceedings of the 11th European Conference on Computer Supported Cooperative Work.: Springer; 2009 Presented at: 11th European Conference on Computer Supported Cooperative Work; September 7-11, 2009; Vienna, Austria p. 63-83. [doi: 10.1007/978-1-84882-854-4_4]

XSL•FO
RenderX

25. Smith BK, Frost J, Albayrak M, Sudhakar R. Integrating glucometers and digital photography as experience capture tools to enhance patient understanding and communication of diabetes self-management practices. Pers Ubiquitous Comput 2006 Aug 15;11(4):273-286. [doi: 10.1007/s00779-006-0087-2]

26. Alsos OA, Das A, Svanæs D. Mobile health IT: the effect of user interface and form factor on doctor-patient communication. Int J Med Inform 2012 Jan;81(1):12-28. [doi: 10.1016/j.ijmedinf.2011.09.004] [Medline: 21996624]

27. Chen Y, Ngo V, Harrison S, Duong V. Unpacking exam-room computing: negotiating computer-use in patient-physician interactions. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '11). New York, NY: ACM; 2011 Presented at: SIGCHI Conference on Human Factors in Computing Systems (CHI '11); May 7-12, 2011; Vancouver, BC p. 3343-3352. [doi: 10.1145/1978942.1979438]

28. Luff P, Heath C. Mobility in collaboration. In: Proceedings of the 1998 ACM Conference on Computer Supported Cooperative Work (CSCW '98). New York, NY: ACM; 1998 Presented at: 1998 ACM Conference on Computer Supported Cooperative Work (CSCW '98); November 14-18, 1998; Seattle, WA p. 305-314. [doi: 10.1145/289444]

29. Lan M, Samy L, Alshurafa N, Suh MK, Ghasemzadeh H, Macabasco-O'Connell A, et al. WANDA: an end-to-end remote health monitoring and analytics system for heart failure patients. In: Proceedings of the Conference on Wireless Health (WH '12). New York, NY: ACM; 2012 Presented at: Wireless Health (WH '12); October 23-25, 2012; San Diego, CA. [doi: 10.1145/2448096.2448105]

30. Radhakrishnan K, Bowles K, Hanlon A, Topaz M, Chittams J. A retrospective study on patient characteristics and telehealth alerts indicative of key medical events for heart failure patients at a home health agency. Telemed J E Health 2013 Sep;19(9):664-670 [FREE Full text] [doi: 10.1089/tmj.2012.0307] [Medline: 23808888]

31. Mamykina L, Mynatt E, Davidson P, Greenblatt D. MAHI: investigation of social scaffolding for reflective thinking in diabetes management. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '08). New York, NY: ACM; 2008 Presented at: SIGCHI Conference on Human Factors in Computing Systems (CHI '08); April 5-10, 2008; Florence, Italy p. 477-486. [doi: 10.1145/1357054.1357131]

32. Tang PC, Overhage JM, Chan AS, Brown NL, Aghighi B, Entwistle MP, et al. Online disease management of diabetes: engaging and motivating patients online with enhanced resources-diabetes (EMPOWER-D), a randomized controlled trial. J Am Med Inform Assoc 2013 May 1;20(3):526-534 [FREE Full text] [doi: 10.1136/amiajnl-2012-001263] [Medline: 23171659]

33. Andersen T, Bjørn P, Kensing F, Moll J. Designing for collaborative interpretation in telemonitoring: re-introducing patients as diagnostic agents. Int J Med Inform 2011 Aug;80(8):e112-e126. [doi: 10.1016/j.ijmedinf.2010.09.010] [Medline: 21067968]

34. Personal Data for the Public Good: New Opportunities to Enrich Understanding of Individual and Population Health.: Health Data Exploration Project; 2014. URL: http://www.rwjf.org/content/dam/farm/reports/reports/2014/rwjf411080 [accessed 2015-01-03] [WebCite Cache ID 6VKABOe6y]

35. Ogden CL, Carroll MD, Kit BK, Flegal KM. Prevalence of Obesity in the United States 2009-2010. NCHS Data Brief, no 82. Hyattsville, MD: National Center for Health Statistics; 2012. URL: http://www.cdc.gov/nchs/data/databriefs/db82.pdf [accessed 2015-01-03] [WebCite Cache ID 6VKBEd5tW]

36. World Health Organization. Geneva, Switzerland: World Health Organization; 2015 Jan. Obesity and overweight. Fact sheet N°311 URL: http://www.who.int/mediacentre/factsheets/fs311/en/ [WebCite Cache ID 6VKIv2eZ0]

37. Centers for Disease Control and Prevention. Atlanta, GA: Centers for Disease Control and Prevention Adult obesity facts URL: http://www.cdc.gov/obesity/data/adult.html [accessed 2015-01-02] [WebCite Cache ID 6VK8dreQA]

38. Richmond R, Kehoe L, Heather N, Wodak A, Webster I. General practitioners' promotion of healthy life styles: what patients think. Aust N Z J Public Health 1996 Apr;20(2):195-200. [Medline: 8799096]

39. Aittasalo M, Miilunpalo S, Kukkonen-Harjula K, Pasanen M. A randomized intervention of physical activity promotion and patient self-monitoring in primary health care. Prev Med 2006 Jan;42(1):40-46. [doi: 10.1016/j.ypmed.2005.10.003] [Medline: 16297442]

40. Elsenbruch S. Abdominal pain in irritable bowel syndrome: a review of putative psychological, neural and neuro-immune mechanisms. Brain Behav Immun 2011 Mar;25(3):386-394. [doi: 10.1016/j.bbi.2010.11.010] [Medline: 21094682]

41. Jamieson AE, Fletcher PC, Schneider MA. Seeking control through the determination of diet: a qualitative investigation of women with irritable bowel syndrome and inflammatory bowel disease. Clin Nurse Spec 2007;21(3):152-160. [doi: 10.1097/01.NUR.0000270015.97457.9c] [Medline: 17495550]

42. Fitbit. URL: https://www.fitbit.com/ca [accessed 2015-08-10] [WebCite Cache ID 6aiAWnPlN]

43. Zia JK, Nguyen L, Heitkemper M, Ladabaum U. The feasibility and usability of a novel mobile phone application, a food and symptom diary, for irritable bowel syndrome patients. Presented at: 17th Neurogastroenterology & Motility Scientific Meeting; September 20-22, 2013; Huntington Beach, CA.

44. Cook J. Chronologic: Tracking Stories of Illness in Chronic Care [dissertation]. Seattle, WA: University of Washington; 2014. URL: https://digital.lib.washington.edu/researchworks/bitstream/handle/1773/26741/Cook_washington_0250O_13283.pdf [accessed 2015-08-11] [WebCite Cache ID 6akt9KNbl]

45. Holtzblatt K, Wendell JB, Wood S. Rapid Contextual Design: A How-to Guide to Key Techniques for User-Centered Design. San Francisco, CA: Morgan Kaufmann Publishers; 2004.

XSL•FO
RenderX

46.   Weight Watchers. URL: https://welcome.weightwatchers.com/ [accessed 2015-08-10] [WebCite Cache ID 6aiCx9cca]

47.   Mohr DC, Cuijpers P, Lehman K. Supportive accountability: a model for providing human support to enhance adherence to eHealth interventions. J Med Internet Res 2011;13(1):e30 [FREE Full text] [doi: 10.2196/jmir.1602] [Medline: 21393123]

48.   Cordeiro F, Epstein DA, Thomaz E, Bales E, Jagannathan AK, Abowd GD, et al. Barriers and negative nudges: exploring challenges in food journaling. In: Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems (CHI '15). New York, NY: ACM; 2015 Presented at: 33rd Annual ACM Conference on Human Factors in Computing Systems (CHI '15); April 18-23, 2015; Seoul, Korea p. 1159-1162. [doi: 10.1145/2702123.2702155]

49.   Tsai CC, Lee G, Raab F, Norman GJ, Sohn T, Griswold WG, et al. Usability and feasibility of PmEB: a mobile phone application for monitoring real time caloric balance. Mobile Netw Appl 2007 Jul 15;12(2-3):173-184. [doi: 10.1007/s11036-007-0014-4]

50.   Bonilla C, Brauer P, Royall D, Keller H, Hanning RM, DiCenso A. Use of electronic dietary assessment tools in primary care: an interdisciplinary perspective. BMC Med Inform Decis Mak 2015;15:14 [FREE Full text] [doi: 10.1186/s12911-015-0138-6] [Medline: 25886381]

51.   van Velsen L, Beaujean DJ, van Gemert-Pijnen JE. Why mobile health app overload drives us crazy, and how to restore the sanity. BMC Med Inform Decis Mak 2013;13:23 [FREE Full text] [doi: 10.1186/1472-6947-13-23] [Medline: 23399513]

52.   Montgomery-Downs HE, Insana SP, Bond JA. Movement toward a novel activity monitoring device. Sleep Breath 2012 Sep;16(3):913-917. [doi: 10.1007/s11325-011-0585-y] [Medline: 21971963]

53.   Grönvall E, Verdezoto N. Beyond self-monitoring: understanding non-functional aspects of home-based healthcare technology. In: Proceedings of the 2013 ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp '13). New York, NY: ACM; 2013 Presented at: 2013 ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp '13); September 8-12, 2013; Zurich, Switzerland p. 587-596. [doi: 10.1145/2493432.2493495]

54.   Paré G, Jaana M, Sicotte C. Systematic review of home telemonitoring for chronic diseases: the evidence base. J Am Med Inform Assoc 2007;14(3):269-277 [FREE Full text] [doi: 10.1197/jamia.M2270] [Medline: 17329725]

55.   Chaudhry SI, Mattera JA, Curtis JP, Spertus JA, Herrin J, Lin Z, et al. Telemonitoring in patients with heart failure. N Engl J Med 2010 Dec 9;363(24):2301-2309 [FREE Full text] [doi: 10.1056/NEJMoa1010029] [Medline: 21080835]

56.   Google Health. URL: http://www.google.com/intl/en-CA/health/about/ [accessed 2015-08-10] [WebCite Cache ID 6aiBctezn]

57.   HealthVault. URL: https://www.healthvault.com/ca/en [accessed 2015-08-10] [WebCite Cache ID 6aiBkytyf]

58.   Landeweerd M, Spil T, Klein R. The Success of Google Search, the Failure of Google Health and the Future of Google Plus. In: Proceedings of the IFIP WG 8.6 International Conference on Transfer and Diffusion of IT (TDIT 2013). Berlin/Heidelberg, Germany: Springer-Verlag; 2013 Presented at: IFIP WG 8.6 International Conference on Transfer and Diffusion of IT (TDIT 2013); June 27-29, 2013; Bangalore, India p. 221-239. [doi: 10.1007/978-3-642-38862-0_14]

59.   Liu LS, Shih PC, Hayes GR. Barriers to the adoption and use of personal health record systems. In: Proceedings of the 2011 iConference (iConference '11). New York, NY: ACM; 2011 Presented at: 2011 iConference (iConference '11); February 8-11, 2011; Seattle, WA p. 363-370. [doi: 10.1145/1940761.1940811]

60.   Microsoft Health. URL: https://www.microsoft.com/microsoft-health/en-us [accessed 2015-08-10] [WebCite Cache ID 6aiEjLbyk]

61.   Google Fit. URL: https://fit.google.com [accessed 2015-08-10] [WebCite Cache ID 6aiEpcCmg]

62.   PatientsLikeMe. URL: https://www.patientslikeme.com/ [accessed 2015-08-09] [WebCite Cache ID 6aiIBpt6W]

63.   Brubaker JR, Lustig C, Hayes GR. PatientsLikeMe: empowerment and representation in a patient-centered social network. In: Proceedings of the 2010 ACM Conference on Computer Supported Cooperative Work. Workshop on Research in Healthcare: Past, Present, and Future. New York, NY: ACM; 2010 Presented at: 2010 ACM Conference on Computer Supported Cooperative Work; February 6-10, 2010; Savannah, GA.

64.   Wicks P, Massagli M, Frost J, Brownstein C, Okun S, Vaughan T, et al. Sharing health data for better outcomes on PatientsLikeMe. J Med Internet Res 2010;12(2):e19 [FREE Full text] [doi: 10.2196/jmir.1549] [Medline: 20542858]

65.   Ramsey A, Lord S, Torrey J, Marsch L, Lardiere M. Paving the way to successful implementation: identifying key barriers to use of technology-based therapeutic tools for behavioral health care. J Behav Health Serv Res 2014 Sep 6:1-17. [doi: 10.1007/s11414-014-9436-5] [Medline: 25192755]

66.   Moukheiber Z. Forbes. 2014 Jun 02. In groundbreaking move For health trackers, Apple teams with Mayo URL: http://www.forbes.com/sites/zinamoukheiber/2014/06/02/in-groundbreaking-move-for-health-trackers-apple-teams-with-mayo-clinic/ [accessed 2015-07-15] [WebCite Cache ID 6a3YWNLZI]

Abbreviations

AHRQ: Agency for Healthcare Research and Quality
API: application programming interface
BMI: body mass index
EHR: electronic health record

FODMAP: fermentable, oligo-, di-, monosaccharides and polyols
IBS: irritable bowel syndrome
NSF: National Science Foundation

XSL•FO
RenderX

Who Uses the Internet as a Source of Nutrition and Dietary Information? An Australian Population Perspective

Christina Mary Pollard[1,2*], PhD, MPH, Grad Dip Diet, BApp Sci; Claire Elizabeth Pulker[1*], MMedSci, BSc (Hons); Xingqiong Meng[3*], PhD, MAppE, MBBS; Deborah Anne Kerr[1*], PhD, MSc, Grad Dip Diet, BApp Sci; Jane Anne Scott[1*], PhD, MPH, Grad Dip Diet, BApp Sci

[1]School of Public Health, Curtin University, Perth, Australia
[2]Department of Health in Western Australia, East Perth, Australia
[3]School of Medicine, Flinders University, Bedford Park, Australia
[*]all authors contributed equally

Corresponding Author:
Christina Mary Pollard, PhD, MPH, Grad Dip Diet, BApp Sci
School of Public Health
Curtin University
Kent Street
GPO Box U1987
Perth, WA 6845
Australia
Phone: 61 892661142
Fax: 61 892667947
Email: C.Pollard@curtin.edu.au

Abstract

Background: The Internet contains a plethora of nutrition information. Health organizations are increasingly using the Internet to deliver population-wide health information and interventions. Effective interventions identify their target population and their needs; however, little is known about use of the Internet as a source of nutrition information.

Objective: The aim was to assess the change in prevalence and demographic characteristics of Western Australian adults accessing the Internet as a source of nutrition information and identify specific information needs.

Methods: Data were pooled from the Western Australian Department of Health's 3-yearly Nutrition Monitoring Survey Series telephone survey between 1995 and 2012 of 7044 participants aged 18 to 64 years. Outcome variables were the main sources of nutrition information used in the last year and yes/no responses to 4 suggestions to what would make it easier to eat a healthy diet. Sociodemographic variables were collected.

Results: The proportion of respondents using the Internet for nutrition information increased from <1% in 1995-2001 to 9.1% in 2004 and 33.7% in 2012. Compared to 2004, logistic regression showed that the odds of using the Internet for this information increased significantly in 2009 (OR 2.84, 95% CI 2.07-3.88) and 2012 (OR 5.20, 95% CI 3.86-7.02, P<.001). Respondents using the Internet as a source were more likely to be female (OR 1.30, 95% CI 1.05-1.60, P=.02), live in a metropolitan area (OR 1.26, 95% CI 1.03-1.54, P=.03), born in countries other than Australia/UK/Ireland (OR 1.41, 95% CI 1.07-1.85, P=.02), more educated (university: OR 2.46, 95% CI 1.77-3.42, P<.001), and were less likely to be older (55-64 years: OR 0.38, 95% CI 0.25-0.57, P<.001). The majority of respondents agreed the following information would assist them to make healthier choices: more ways to prepare healthy foods (72.0%, 95% CI 70.7-73.3), quicker ways to prepare healthy foods (79.0%, 95% CI 77.8-80.1), how to choose healthy foods (68.8%, 95% CI 67.5-70.1), and knowing more about cooking (54.7%, 95% CI 53.3-56.1). Those using the Internet for nutrition information were more likely than nonusers to want to know quicker ways to prepare healthy foods (83.0% vs 78.1%, P=.005) and information on choosing healthy foods (76.3% vs 67.3%, P<.001).

Conclusions: Use of the Internet as a main source of nutrition information has grown rapidly since 2004; one-third of Western Australian adults reported using the Internet for this purpose in 2012. Information on preparing healthy foods (ideas, quicker ways), choosing ingredients, and knowing more about cooking would make it easier to eat a healthy diet. For Internet users, emphasis should be on quicker ways and choosing ingredients. These finding have implications for policy makers and practitioners and suggest that traditional health promotion tactics should continue to be used to reach the broader population.

XSL•FO
RenderX

Introduction

Use of technology for communication has increased and traditional sources of information have changed in importance over time. Use of the Internet in Australian homes has risen dramatically from 3.4% of households in 1996 to 80% in 2012 [1,2]. Searching for health information is the third most popular use of the Internet [3]. Recent research in the United States, Norway, and Canada estimates that almost half of Internet users search for information on diet and nutrition [4-7]. The Norwegian research estimated that 40% of online health seekers are willing to change their diet or other lifestyle aspects [5]. The size and scale of the Internet raises concerns about the credibility and reliability of the nutrition and dietary information available [8].

In Australia, diet is a significant factor contributing to the burden of disease [9], accounting for 43% of the total risk factor burden [10] at an estimated cost of AUS $67 billion per year in 2010 [11]. The general population is bombarded with misleading or inaccurate nutrition messages [12]. Population-wide health interventions aim to improve the quality and quantity of dietary choices by providing sound nutrition and dietary information. Interventions that provide accurate and balanced information enable the target population to make well-informed, healthy choices [13]. The World Health Organization encourages countries to develop food-based dietary guidelines and most nutrition interventions are based on these [14]. Australian Dietary Guidelines have provided evidence-based recommendations for a healthy diet over several decades [15-18]. The challenge for health authorities is to promote changes in dietary intake, replacing eating patterns predominated by energy-dense nutrient-poor "discretionary" foods with eating patterns that are in-line with the dietary guidelines [19].

Online health interventions have the capacity to influence voluntary behavior change and have the advantage of lower costs and increased reach when compared with more traditional channels, such as print media, with similar impact [20]. In addition, the Internet offers the ability to individually tailor intervention messages or provide interactivity that is not available with other channels [21]. Web 2.0 features have been demonstrated to have great potential for Internet-based health interventions [21]. Social media apps utilizing Web 2.0 allow for the creation and exchange of user-generated content [22] that moves beyond one-directional read-only website content [23]. More than half of US adults use social media [24] and two-thirds of Australian adults who used the Internet at home went online for social networking [2]. A third of online health seekers in the United States used social media to source health-related information in 2008 [25]. Evidence suggests that people using social media for health information are more likely to read rather than create or contribute [26]. The fastest growing source of information about health and nutrition for younger adults is social networking sites such as YouTube, Facebook, and Twitter [27].

As use of the Internet has grown, the potential to reach a large number of people has made it appealing for dissemination of nutrition interventions. Government and not-for-profit health organizations increasingly use the Internet for population-wide health interventions [20,21,28]. Effective nutrition interventions require use of timely and appropriate delivery channels for dissemination. An understanding of the population and their current media use assists with selecting appropriate social marketing communication channels [29]. At the time of conducting this research, little was known about use of the Internet as a source of nutrition and dietary information in Australia and the types of information sought.

The primary aim of this paper was to assess the change in prevalence and demographic characteristics of those using the Internet as a source to obtain nutrition and dietary information among Western Australian adults over 2 decades. To assist with the development of population-wide nutrition interventions, the association between use of the Internet as a source of nutrition and dietary information and the respondents' perception of what would make it easier for them to eat a healthy diet was explored.

Methods

Sample

The data were part of the Department of Health in Western Australia's Nutrition Monitoring Survey Series (NMSS), which aims to investigate the nutrition knowledge, attitudes, and behaviors of Western Australian adults related to the Dietary Guidelines to prioritize and monitor the impact of nutrition interventions. Computer-assisted telephone interviews with 7044 Western Australian adults aged 18 to 64 years were conducted from July to August in 1995, 1998, 2001, 2004, 2009, and 2012. The 1995 sample was stratified by geographic area and the 1998, 2001, and 2004 samples were quota sampled by gender and geographic area. The telephone numbers were selected randomly by a computer-generated random digit dialing program. In 2009 and 2012, the samples were randomly drawn from the 2008 and 2011 Electronic White Pages for Western Australia, respectively, and stratified according to area of residence. The details of survey and sampling strategies can be found elsewhere [30].

Outcome Measurements

The main purpose of this study was to assess change in prevalence and demographic characteristics of Western Australian adults using the Internet as a source of nutrition and dietary information. To answer the main purpose, respondents were asked an open-ended question: "In the last 12 months, what have been your main sources of nutrition and dietary information?" After first responses were given, the interviewer

probed "anything else?" The main sources of nutrition information used were television advertising and programs, magazine articles, books, and the Internet; this study reports on respondents' use of the Internet. We compared respondents who used the Internet as the main source of nutrition and dietary information with those who did not mention using the Internet. The secondary aim of the study was to explore the association between use of the Internet as a source of nutrition and dietary information and the respondents' perception of what would make it easier for them to eat a healthy diet to assist with the development of population-wide nutrition interventions. To answer the secondary aim, respondents were asked "Which of the following would make it easier for you or your family to eat a healthy diet?" with the following options: (1) knowing more ways of preparing healthy foods, (2) knowing quicker ways of preparing healthy foods, (3) having more information to help me decide if foods are healthy, and (4) knowing more about cooking (all answered "yes" or "no"). The response options for making it easier to eat a healthy diet were identified during the development stages for the NMSS tool in which open-ended questions were used and typical responses identified. Participants' sociodemographic information (gender, age, education, area of residence, household income, employment status, country of birth, area of residence, height, and weight) was also collected.

Statistical Analysis

The data were collected to be representative of the Western Australian population. Data for all the years were pooled and weighted to account for sampling design and adjusted for age, sex, and geographic area to a single standard population to allow for comparisons over time. The standard population used was the 2011 estimated resident population of Western Australia [31] because it was the most recent census year.

The descriptive statistics report the prevalence and 95% confidence interval (95% CI) of using the Internet to obtain nutrition and dietary information by survey year, gender, and age. Binary logistic regression was used to assess the association between respondents using the Internet as a source of nutrition and dietary information and sociodemographic characteristics. The covariates in the full model included survey year as a measure of time, gender, age, education, household income, employment status, country of birth, area of residence (metropolitan vs nonmetropolitan), body mass index (BMI) categories, and Socioeconomic Indexes for Areas (SEIFA) quintile [32]. Variables with P value <.05 were retained in the final model and reported. Data from surveys administered in 2009 and 2012 were compared for Internet use by gender and age to provide a more recent analysis.

All analyses were performed using Stata 12.0 (StataCorp LP, College Station, TX, USA) under the survey module and a P value <.05 was regarded as statistically significant.

Results

Between 1995 and 2012, a total of 7044 adults participated in the NMSS. Details of the sample and demographic information are shown in Table 1.

The main purpose of this study was to assess change in prevalence and demographic characteristics of Western Australian adults using the Internet as a source of nutrition and dietary information. Table 2 shows that there was a significant increase in the proportion of the population using the Internet as a source of nutrition and dietary information between 1995 (0.2%, 95% CI 0.0-1.2), 2004 (9.1%, 95% CI 7.3-11.4), and 2012 (33.7%, 95% CI 30.8-36.8, P<.001). Across all surveys, more females (19.6%, 95% CI 17.9-21.5) used the Internet than males (14.2%, 95% CI 12.5-16.1, P=.005) and no significant difference was found by age. The top 5 main sources of nutrition information in 2012 were the Internet (33.7%, 95% CI 30.8-36.8), television programs (23.7%, 95% CI 21.2-26.2), magazine articles (22.5%, 95% CI 20.1-25.0), television advertising (22.0%, 95% CI 19.6-24.6), and books (19.0%, 95% CI 16.7-21.5).

Table 1. Sample demographics of the Nutrition Monitoring Survey Series, Western Australia, 1995-2012.

Demographic characteristics	Survey year, n						Total, %[a] N=7044
	1995 n=1002	1998 n=1004	2001 n=1004	2004 n=1202	2009 n=1284	2012 n=1548	
Sex							
Female	631	502	502	601	830	1005	49.18
Male	371	502	502	601	454	543	50.82
Age group (years)							
18-24	119	110	118	103	71	66	15.80
25-34	257	210	245	232	180	144	23.01
35-44	291	305	296	333	340	377	22.41
45-54	207	234	212	297	356	466	21.33
55-64	128	145	133	237	337	495	17.45
Area of residence							
Metropolitan	748	751	75	601	965	1011	78.33
Remote areas	51	63	62	150	29	82	4.80
Rural areas	203	190	18	451	290	455	16.87
Education							
Less than high school	376	336	303	330	221	211	19.67
High school	251	237	265	257	178	198	21.98
Trade/certificate/diploma	90	95	77	177	481	632	25.47
University degree	284	336	344	435	399	504	33.58
Missing	1	0	1	3	5	3	0.30
Household income (AUS $)							
≤$60,000	748	603	558	603	349	346	37.47
>$60,000	174	305	340	560	814	1024	51.40
Missing	80	96	106	39	121	178	11.13
Employment status							
Currently not in paid employment	330	263	278	285	364	408	26.82
Currently in paid employment	669	741	726	917	920	1139	73.14
Missing	3	0	0	0	0	1	0.04
Country of birth							
Australia	656	665	668	868	867	1122	68.64
UK/Ireland	189	209	193	155	202	221	15.96
Other countries	157	130	143	179	214	205	16.38
Missing	0	0	0	0	1	0	0.02

[a] Percentages were weighted for probability of sample selection and adjusted by age, sex, and geographic area to the 2011 Estimated Resident Population of Western Australia.

Table 2. Prevalence of using the Internet as a source to obtain nutrition and dietary information, the Nutrition Monitoring Survey Series, Western Australia, 1995-2012.

Variable	No, % (95% CI)[a]	Yes, % (95% CI)[a]	P [b]
Survey year			<.001
1995	99.83 (98.84, 99.98)	0.17 (0.02, 1.16)	
1998	99.43 (98.78, 99.73)	0.57 (0.27, 1.22)	
2001	99.84 (98.89, 99.98)	0.16 (0.02, 1.11)	
2004	90.86 (88.62, 92.70)	9.14 (7.30, 11.38)	
2009	77.09 (73.95, 79.96)	22.91 (20.04, 26.05)	
2012	66.27 (63.23, 69.18)	33.73 (30.82, 36.77)	
Gender			<.001
Female	80.36 (78.51, 82.09)	19.64 (17.91, 21.49)	
Male	85.80 (83.95, 87.47)	14.20 (12.53, 16.05)	
Age group (years)			.21
18-24	79.92 (74.69, 84.30)	20.08 (15.70, 25.31)	
25-34	83.93 (81.01, 86.47)	16.07 (13.53, 18.99)	
35-44	83.22 (81.04, 85.19)	16.78 (14.81, 18.96)	
45-54	82.89 (80.51, 85.04)	17.11 (14.96, 19.49)	
55-64	85.14 (82.70, 87.29)	14.86 (12.71, 17.30)	
Total	83.13 (81.83, 84.35)	16.87 (15.65, 18.17)	

[a] Percentages were weighted for probability of sample selection and adjusted by age, sex, and geographic area to the 2011 Estimated Resident Population of Western Australia.

[b] P values were derived from survey design–based Pearson chi-square test.

The logistic regression results showed a sharp increase in using the Internet for obtaining nutrition and dietary information from 2004 (Table 3 and Figure 1). After adjustment for the model covariates, gender showed a difference in the odds of using the Internet as a source. Females were more likely to use the Internet as a nutrition information source than males (OR 1.30, 95% CI 1.05-1.60, P=.02). Respondents living in metropolitan areas (OR 1.26, 95% CI 1.03-1.54, P=.03); born in countries other than Australia, United Kingdom, or Ireland (OR 1.41, 95% CI 1.07-1.85, P=.02); and with more than high school level education (tertiary educated: OR 2.46, 95% CI 1.77-3.42, P<.001) were more likely to use the Internet as a source (Table 3). Respondents aged between 35 and 64 years were less likely to use the Internet as a source compared to those aged between 18 and 24 years (P<.001) (Table 3 and Figure 2).

Table 3. Factors associated with using the Internet as a source of obtaining nutrition and dietary information, the Nutrition Monitoring Survey Series, Western Australia, 1995-2012.

Factor	OR (95% CI)	P [a]
Survey year		<.001
1995	0.02 (0, 0.12)	
1998	0.06 (0.03, 0.13)	
2001	0.02 (0.00, 0.11)	
2004	1.00	
2009	2.84 (2.07, 3.88)	
2012	5.20 (3.86, 7.02)	
Gender		.02
Male	1.00	
Female	1.30 (1.05, 1.60)	
Age group (years)		<.001
18-24	1.00	
25-34	0.78 (0.51, 1.20)	
35-44	0.64 (0.43, 0.95)	
45-54	0.51 (0.34, 0.76)	
55-64	0.38 (0.25, 0.57)	
Education attainment		<.001
Less than high school	1.00	
High school	1.16 (0.76, 1.76)	
TAFE/trade/diploma	1.80 (1.29, 2.52)	
Tertiary	2.46 (1.77, 3.42)	
Country of birth		.02
Australia/UK/Ireland	1.00	
Other countries	1.41 (1.07, 1.85)	
Residential area		.03
Nonmetropolitan	1.00	
Metropolitan	1.26 (1.03, 1.54)	

[a] Results were derived from a binary logistic regression under survey module. *P* values were derived from Wald test.

The secondary aim of this study was to explore the association between use of the Internet as a source of nutrition and dietary information and the respondents' perception of what would make it easier for them to eat a healthy diet to assist with the development of population-wide nutrition interventions. A high proportion of all respondents agreed that the following information would make it easier for them to eat a healthier diet: knowing more ways of preparing healthy foods (71.99%, 95% CI 70.67-73.27), knowing quicker ways of preparing healthy foods (78.96%, 95% CI 77.80-80.08), more information to help decide if foods are healthy (68.80%, 95% CI 67.49-70.07), and knowing more about cooking (54.68%, 95% CI 53.25-56.11). Table 4 shows that when compared to respondents who did not use the Internet to obtain nutrition and dietary information, a statistically significantly greater proportion of people who used the Internet agreed that knowing quicker ways of preparing healthy foods (83.03% vs 78.14%, *P*=.005) and having more information to help decide if foods are healthy (76.28% vs 67.27%, *P*<.001) would make it easier for them to eat a healthier diet.

When comparing survey results for 2009 and 2012, Table 5 shows that a significantly higher percentage of females aged between 25 and 54 years used the Internet as a source of nutrition and dietary information in 2012 with the highest users being females aged between 25 and 34 years (58.8%, 95% CI 46.4-70.1, *P*<.001). There were also higher numbers of males aged between 35 and 44 years using the Internet as a source of nutrition and dietary information in 2012 compared to 2009 (36.6%, 95% CI 27.1-47.2 vs 17.0%, 95% CI 10.8-25.8, *P*=.002).

XSL•FO
RenderX

Table 4. Association between using the Internet as a source to obtain nutrition and dietary information and perception of whether it would be easier for respondents to eat healthy diet.

Would make easier to eat healthy diet	Participants, % (95% CI)[a]		P [b]	Total, % (95% CI)[a]
	Not using Internet	Using Internet		
Knowing more ways of preparing healthy foods (n=7007)			.06	
No	28.65 (27.29, 30.05)	24.88 (21.45, 28.65)		28.01 (26.73, 29.33)
Yes	71.35 (69.95, 72.71)	75.12 (71.35, 78.55)		71.99 (70.67, 73.27)
Knowing quicker ways of preparing healthy foods (n=7019)			.005	
No	21.86 (20.65, 23.12)	16.97 (14.27, 20.07)		21.04 (19.92, 22.20)
Yes	78.14 (76.88, 79.35)	83.03 (79.93, 85.73)		78.96 (77.80, 80.08)
More information on healthy foods (n=6996)			<.001	
No	32.73 (31.35, 34.15)	23.72 (20.58, 27.18)		31.20 (29.93, 32.51)
Yes	67.27 (65.85, 68.65)	76.28 (72.82, 79.42)		68.80 (67.49, 70.07)
If I knew more about cooking (n=7011)			.33	
No	45.69 (44.19, 47.20)	43.49 (39.42, 47.64)		45.32 (43.89, 46.75)
Yes	54.31 (52.80, 55.81)	56.51 (52.36, 60.58)		54.68 (53.25, 56.11)

[a] Percentages were weighted for probability of sample selection and adjusted by age, sex, and geographic area to the 2011 Estimated Resident Population of Western Australia.

[b] P values were derived from survey design–based Pearson chi-square test.

Table 5. Prevalence of using Internet as a source obtaining nutrition and dietary information, the Nutrition Monitoring Survey Series, Western Australia, 2009 and 2012.

Age range	2009 Yes, % (95% CI)[a]		P	2012 Yes, % (95% CI)[a]		P
	Female	Male		Female	Male	
18-24 years	31.58 (18.15, 49.01)	34.56 (19.05, 54.23)	.81	56.81 (38.09, 73.77)	25.80 (13.59, 43.45)	.01
25-34 years	28.40 (20.12, 38.44)	32.83 (21.07, 47.22)	.59	58.77 (46.40, 70.13)	32.67 (20.73, 47.38)	.005
35-44 years	24.19 (18.99, 30.27)	16.98 (10.76, 25.76)	.13	39.66 (33.37, 46.31)	36.57 (27.08, 47.24)	.62
45-54 years	16.91 (12.42, 22.62)	18.21 (11.58, 27.45)	.79	33.19 (27.41, 39.52)	27.11 (20.14, 35.43)	.22
55-64 years	13.48 (9.14, 19.45)	18.21 (11.64, 27.34)	0.32	23.39 (18.58, 29.01)	22.39 (16.44, 29.72)	.82

[a] Percentages were weighted for probability of sample selection and adjusted by age, sex, and geographic area to the 2011 Estimated Resident Population of Western Australia.

Figure 1. Proportion of participants using the Internet as a source of obtaining nutrition and dietary information over the survey period, the Nutrition Monitoring Survey Series, Western Australia, 1995-2012 (derived after the logistic regression).
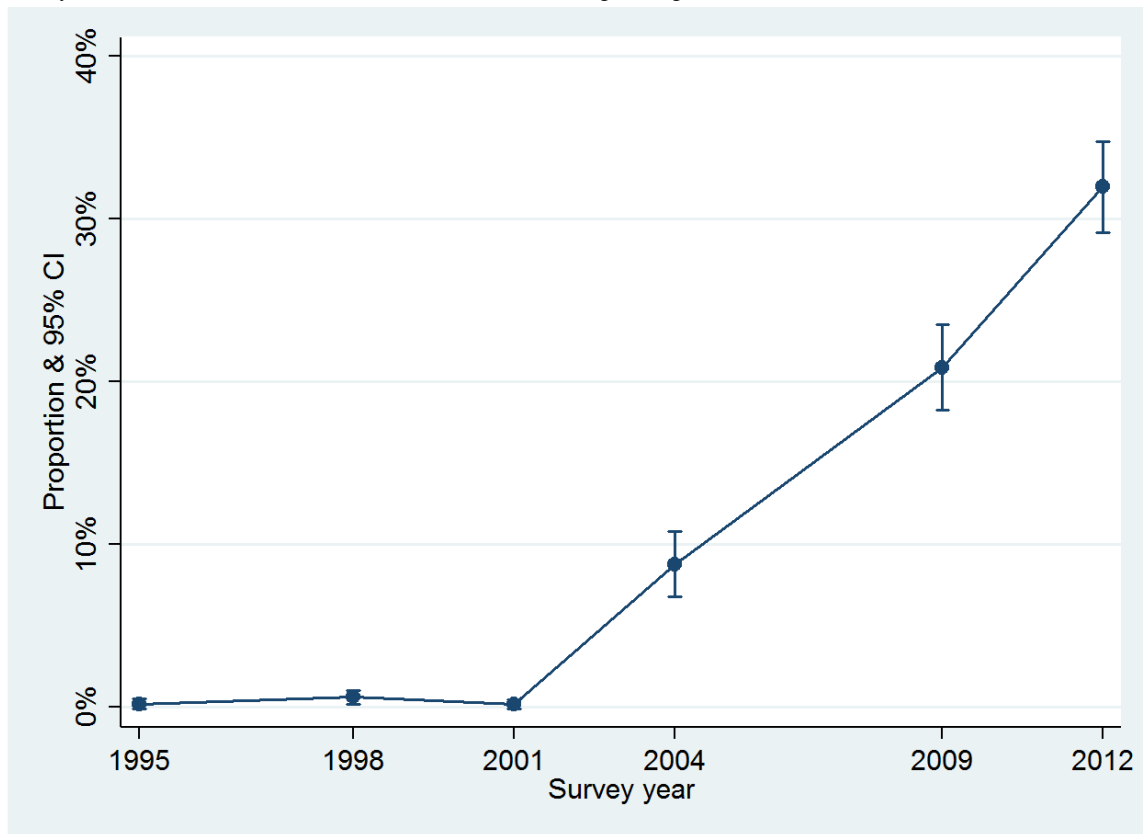
Figure 2. Proportion of respondents using the Internet as a source of nutrition and dietary information over the survey period by gender and age, the Nutrition Monitoring Survey Series, Western Australia, 1995-2012 (derived after logistic regression).
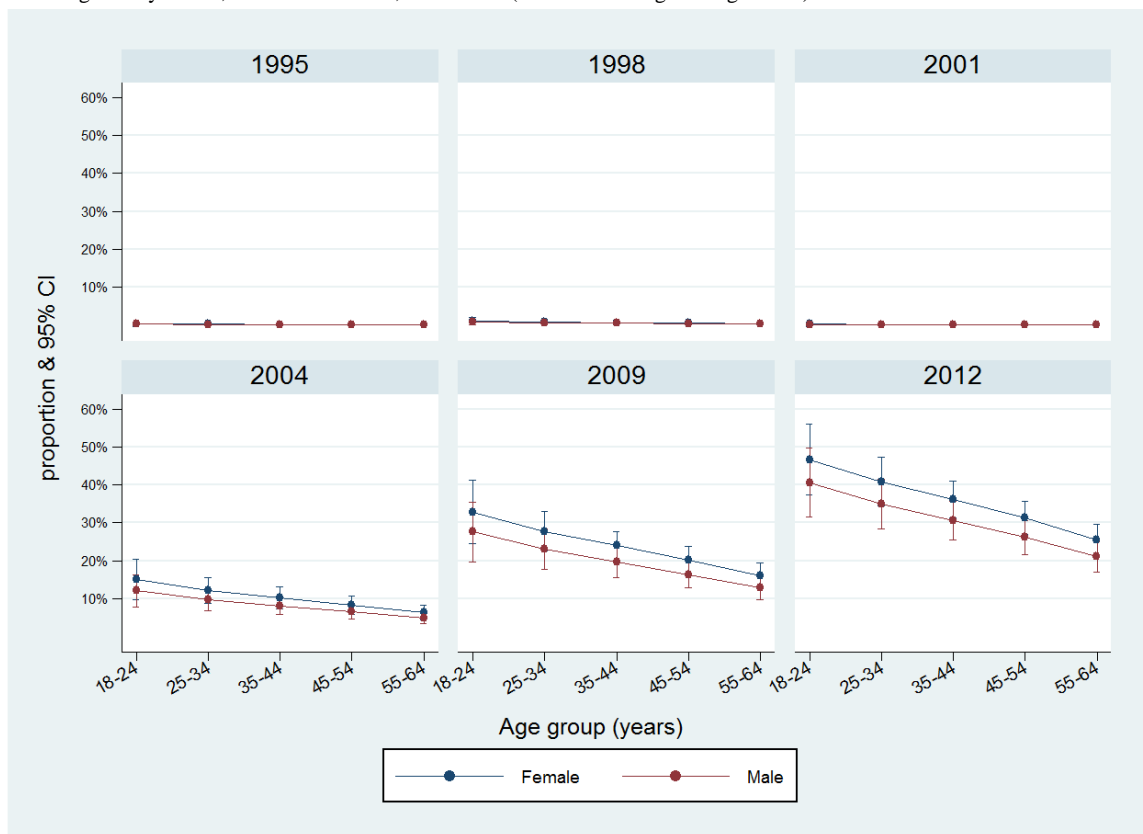
Discussion

Prevalence and Characteristics

This study is unique because it examines characteristics of Australian adults' use of the Internet for nutrition and dietary information and changes in usage over 2 decades. The study found dramatic growth in use of the Internet as a main source of nutrition and dietary information from 9.1% in 2004 to 33.7% in 2012 when it became the most popular main source of nutrition information. This is consistent with overall growth in Internet use in Australian homes, which increased from 3.4% in 1996 to 56% in 2004 and to 83% in 2012 [1,2,33]. However, only one-third of the Western Australian population reported using the Internet as a source of nutrition and dietary information in 2012, which was a smaller proportion than expected. A similar study in Canada found that 51% of the adult population reported using the Internet for nutrition information in 2008 [6]. The proportion of US adults using the Internet to source diet, weight, and physical activity information was 42.8% in 2011 [7]. The Pew Internet & American Life project found that 49% of Internet users had searched for information about diet and nutrition in 2006 [4] (which is equivalent to 34% of the US population [34]). There is limited Australian population research regarding sources of nutrition and dietary information. A 1999 South Australian study of 603 adults did not reveal the Internet as a commonly used source of nutrition information [35]; however, a Queensland convenience sample of 94 adults in 2013 reported that the Internet (63%) was the most utilized source of nutrition information [36]. These findings may reflect cultural differences in requirements for nutrition information. Further research is recommended to understand the implications of the differences in use of the Internet as a resource for nutrition and dietary information.

Comparison of the most recent data available from Australia, the United States, and Canada shows that a similar proportion of the adult population use the Internet overall. In Australia, 83% of adults used the Internet at home in 2012 [2], 81% of US adults used the Internet in 2012 [34], and 83% of Canadian adults used the Internet at home in 2012 [37]. In addition, most Western Australian households reported using the Internet every day (85%) and using a high-speed broadband connection (79%) [2]. A similar proportion of US households accessed the Internet at home using broadband (70%) [38,39] and almost all Canadian Internet users (97%) connected via high-speed broadband [37]. Recent statistics show 80% of US adult Internet users [3] and 70% of Canadian adult Internet users [40] searched for health information online. Improvements in search engine technology have been a key factor in increased online health information seeking over the past decade [41]. Given the rapid increase in use of the Internet in recent years in this study, it is important that governments continue to monitor the prevalence of searching for nutrition and dietary information online.

Western Australian adults using the Internet as a source of nutrition and dietary information were significantly more likely to be female; living in a metropolitan area; born in countries other than Australia, the United Kingdom, and Ireland; more educated; and younger. This is consistent with other studies

from the United States, Switzerland, and Canada, which found that adults using the Internet as a source of dietary information are more likely to be female [7,35], more educated [6,7,35], and younger [6,7,35]. Canadian adults who earned a higher income were more likely to use the Internet for nutrition information [6]; however, in this study, household income was not associated with using the Internet for this purpose. Further, online seekers of general health information from the United States, France, and Germany were more likely to be of higher social class [36,42], women [42,43], and more educated [43]. Our findings reinforce the importance of sociodemographic differences when targeting and developing nutrition interventions, particularly gender and education. Reasons for these differences and implications for engagement need further exploration.

Weight status was not associated with using the Internet for nutrition and dietary information in this study. This is surprising because the Internet has been identified as an important source of information and support for individuals with stigmatized health conditions [44,45], including obesity [46]. American adults using the Internet to help with diet, weight, and physical activity were more likely to have a higher BMI [7] and Canadian adults using the Internet for nutrition information were more likely to have tried a popular diet in the past year [6]. Obese adults from Australia report using the Internet to search for information to assess their likelihood of developing a serious disease as a result of their obesity and ways to minimize the health risks, including healthy recipes and ideas to increase physical activity [45]. Use of the Internet as a source of nutrition and dietary information by obese adults remains an important area to monitor in Australia to provide input to future public health interventions.

Internet-Based Information That Would Make it Easier to Eat Healthily

Western Australians who used the Internet as a source of nutrition and dietary information were more likely than nonusers to agree that quicker ways of preparing healthy foods and help deciding if foods are healthy would help them to eat more healthily. The majority of all respondents also agreed that information about preparing healthy foods and cooking would assist. These findings are consistent with identified barriers to healthy eating. Barriers typically include time pressures, the desire for convenience, and lack of motivation to cook, rather than a lack of skills or knowledge [47]. The US Food Hero social media nutrition intervention, which focused on the importance of serving healthy meals, aimed to provide healthy recipes that overcame time and cost barriers [48]. The US Supplemental Nutrition Assistance Program (SNAP) participants reported that the activities that would motivate them to continue to use an Internet resource for food and nutrition included recipe or product ratings, blogs, and discussion boards [49].

Recent Survey Results

Data from the 2009 and 2012 surveys were examined to explore recent changes in prevalence and the demographic profile of Internet users to guide current nutrition program development. Results showed that more than half of females aged between 18 and 34 years used the Internet as a source of nutrition and dietary information in 2012. The percentage of female users

aged between 25 and 34 years doubled between 2009 and 2012 as did the percentage of male users aged between 35 and 44 years (to 37%). These results suggest that the Internet is increasing in importance for males and females. The opposite trend was seen for males aged between 18 and 24 years. One possibility is that younger males, typically the early adopters of newer technologies, are using other sources of information (eg, smartphone apps) and this should be investigated in further research. The demographic characteristics of US Internet users were also found to change significantly between 2007 and 2011 [7].

The recent increase in use of the Internet for nutrition and dietary information by younger females aged between 18 and 34 years and males aged between 35 and 44 years shows the dynamic nature of the resource. Although the overall proportion of Western Australians using the Internet as a source was relatively lower than results of similar studies, there are subgroups emerging that rely on the Internet more heavily. Changes in the demographic characteristics of users over time is an important area to continue to monitor.

Quality of Information

The NMSS is a government survey that contributes evidence of prevalence of use and demographic information to Western Australian policy and programs. It is critical that high-quality information is used to guide population dietary decision making. The size and scale of the Internet raises concerns about the credibility and reliability of the nutrition and dietary information available [8]. Most people use general search engines to find health information online [43] that present search results based on a page ranking system that is open to manipulation through website design [41]. Analysis of selected Canadian websites commonly used for nutrition information found that commercial websites typically contained some poor and misleading advice [50]. The Internet information most likely to be viewed for weight loss and weight management was also found to contain inaccurate information [41]. More accurate information was provided by medical, government, and university websites; however, they appeared in the second and third pages of the website searches so were less likely to be viewed [41]. Some research suggests that Internet users do not discriminate much in terms of quality of information for nutrition information [4,6,50,51]. A large proportion of online health seekers do not consistently check the source and date of the health information they find online [4] and use commercial websites to seek health and nutrition information [50]. Less than half of Canadians rated the Internet as a source of nutrition information that was very or extremely credible [6]. Recognition of health brands that are trusted in the non-Internet world have been shown to assist with identifying which Internet information to trust [51]. Health organizations need to ensure quality nutrition and dietary information is widely available and accessible on the Internet, which could include introducing certification of websites that provide trustworthy information.

Use of the Internet for Population-Wide Nutrition Interventions

Use of the Internet as a source of nutrition and dietary information has shown recent rapid uptake by Western Australian adults making it appealing for dissemination of public health interventions. Internet-based nutrition and dietary interventions should incorporate good practice characteristics [52] including identifying and understanding the target audience [29] and then customizing and tailoring communications to meet their needs [53]. For Internet-based interventions to be effective, they need to address issues including levels of participation and adherence [22], and fully utilize the interactive nature of Web 2.0 social media platforms [23]. Organizations that choose to engage in social media should maintain content and participate regularly [29], monitor poor quality health information, and provide the credible alternative [23], which can be time consuming to manage [54]. The increasing use of the Internet by health organizations for population-wide health interventions highlights the need for guidelines for effective communication strategies, such as the toolkits from the Centers for Disease Control and Prevention in the United States [28].

In this study, the majority of the population did not report that they used the Internet as a main source of nutrition and dietary information. It is important for population-wide education, such as nutrition interventions, to be inclusive and reach the majority of the population, in this case, reaching adults that report using sources other than the Internet. Australian adults living in regional areas, older than 65 years, living alone, or with low household income are less likely to use the Internet in general [55]. Use of multiple channels as appropriate to the intervention and target population is advised [56]. Messages are reinforced when communicated across a number of channels; for instance, integrating traditional public health intervention tactics with social media [53].

Another important consideration for Internet-based nutrition and dietary interventions is the level of literacy required. Almost half of Australians aged between 15 and 74 years have literacy skills below the level deemed suitable to meet societal demands, including using the Internet [57]. For Internet-based interventions to be inclusive, they should be designed to meet the needs of those with limited literacy [23] by introducing non-text-based social media, including use of images, illustrations, video, and sound [58]. Use of traditional health intervention tactics to engage with groups identified as hard to reach via the Internet can remain effective. For example, an Australian telephone-based nutrition intervention has successfully recruited participants from more disadvantaged and regional areas [59,60]. The intervention recently incorporated an interactive website to assist with recruiting new participants and engaging existing participants [61].

Limitations

There are strengths and limitations to the current study. The results can be generalized to Western Australia due to the high response rate and representative sample selection. Development of survey question terminology including "the Internet" was completed in 1995. Due to the recent rapid uptake of the Internet and other developments, including Web 2.0 social media platforms and smartphone technology, terminology may need to be updated in future surveys to reflect commonly used language to ensure usage is captured fully. Use of the Internet to source nutrition and dietary information has been reported

for the Western Australian adult population, not of Western Australian Internet users. In future, collecting data on use of the Internet, device used, and other relevant technology may help to provide context to use of the Internet to source health and nutrition information. Frequency of use of the Internet to source nutrition and dietary information, the websites that were used, and quality of information sourced were not measured and more information is urgently needed in this area. The study does not specifically investigate the information sources used depending on the information required; for example, information sources for how to feed a toddler a healthy diet may differ to seeking advice on losing weight or healthy recipe ideas for family eating. The cross-sectional survey results are self-reported and are not validated against objective measures; however, it provides useful evidence to measure public attitudes. Height and weight measures are self-reported; however, the use of a correction formula attempted to account for possible underestimation of weight status. This is a population study but there may be differences between Western Australia and other Australian states and territories; for example, more Western Australian households are connected to broadband Internet using mobile broadband (41% vs national average of 33%) [2], which could lead to differences in use of the Internet for nutrition and dietary information. Further research is needed to explore population use of the Internet for nutrition and dietary information and how to deliver public health interventions effectively.

Conclusions

This study found that there had been dramatic growth in using the Internet as a source of nutrition and dietary information since 2004; however, the majority of the adult population still obtain their information from other sources. Relatively fewer Western Australians used the Internet for this purpose when compared with other Western countries, but their demographic characteristics were broadly consistent. This study found that increased weight was not associated with use of the Internet as a source, which was surprising because the Internet has been identified as an important resource for individuals with stigmatized health conditions, including obesity. Given the rapid increase in use of the Internet in recent years, it is likely that prevalence of using the Internet to source nutrition and dietary information will continue to change. Changes in the prevalence and characteristics of users over time are important areas to continue to monitor to inform future development of nutrition interventions.

The Internet provides a cost-effective platform to reach the identified users with nutrition and dietary interventions, but should be integrated with traditional health promotion tactics to reach the broader population. Policy makers and practitioners delivering Internet-based nutrition interventions should ensure they identify and understand the target population, and customize and tailor communications to meet their needs. Use of non–text-based social media, including images, illustrations, video, and sound, should be included to meet the needs of those with limited literacy. Provision of credible, reliable, and practical information is recommended, including quicker ways to prepare healthy foods and how to choose healthy foods. It is also important for policy makers to improve provision of quality nutrition and dietary information on the Internet generally, which could include certification of websites that provide trustworthy information.

Conflicts of Interest

None declared.

References

1. Australian Bureau of Statistics. Use of the Internet by Householders, Australia (cat no 8147). Canberra: ABS; 1998. URL: http://www.ausstats.abs.gov.au/ausstats/subscriber.nsf/0/CA25687100069892CA25688900206390/$File/81470_Feb%201998.pdf [accessed 2015-03-09] [WebCite Cache ID 6Wta9n3pX]
2. Australian Bureau of Statistics. Household Use of Information Technology, Australia, 2012-2013 (cat no 8146). Canberra: ABS; 2014. URL: http://www.abs.gov.au/ausstats/abs@.nsf/Lookup/8146.0Chapter12012-13 [accessed 2015-03-09] [WebCite Cache ID 6WtZuW6Dm]
3. Fox S. Health Topics. Washington, DC: Pew Internet & American Life Project; 2011 Feb 01. URL: http://www.pewinternet.org/files/old-media//Files/Reports/2011/PIP_Health_Topics.pdf [accessed 2015-03-10] [WebCite Cache ID 6WuhdBEU9]
4. Fox S. Online Health Search 2006. Washington, DC: Pew Internet & American Life Project; 2006 Oct 29. URL: http://www.pewinternet.org/files/old-media//Files/Reports/2006/PIP_Online_Health_2006.pdf.pdf [accessed 2015-03-10] [WebCite Cache ID 6WuhuO4XP]
5. Wangberg S, Andreassen H, Kummervold P, Wynn R, Sørensen T. Use of the internet for health purposes: trends in Norway 2000-2010. Scand J Caring Sci 2009 Dec;23(4):691-696. [doi: 10.1111/j.1471-6712.2008.00662.x] [Medline: 19473317]

XSL•FO
RenderX

6. Goodman S, Hammond D, Pillo-Blocka F, Glanville T, Jenkins R. Use of nutritional information in Canada: national trends between 2004 and 2008. J Nutr Educ Behav 2011;43(5):356-365. [doi: 10.1016/j.jneb.2011.02.008] [Medline: 21906548]

7. McCully SN, Don BP, Updegraff JA. Using the Internet to help with diet, weight, and physical activity: results from the Health Information National Trends Survey (HINTS). J Med Internet Res 2013;15(8):e148 [FREE Full text] [doi: 10.2196/jmir.2612] [Medline: 23906945]

8. Brug J, Oenema A, Kroeze W, Raat H. The internet and nutrition education: challenges and opportunities. Eur J Clin Nutr 2005 Aug;59 Suppl 1:S130-7; discussion S138. [doi: 10.1038/sj.ejcn.1602186] [Medline: 16052181]

9. Australian National Preventive Health Agency. State of Preventive Health 2013. Report to the Australian Government Minister for Health. Canberra: ANPHA; 2013. URL: http://www.anpha.gov.au/internet/anpha/publishing.nsf/Content/ state-of-prev-health-2013/$FILE/ANPHA%2041087%20State%20of%20Preventative%20Health_acc_pdf.pdf [accessed 2015-03-08] [WebCite Cache ID 6WtW49sjb]

10. Binns C, Howat P, Jancey J. Health promotion success in Australia and a note of warning. Health Promot J Austr 2014 Dec;25(3):157-159. [doi: 10.1071/HEv25n3_ED] [Medline: 25487688]

11. Australian Institute of Health and Welfare. Australia's Health 2010. Canberra: Australian Institute of Health and Welfare; 2010. URL: http://www.aihw.gov.au/WorkArea/DownloadAsset.aspx?id=6442452962 [accessed 2015-03-10] [WebCite Cache ID 6Wuen6q5b]

12. Ayoob K, Duyff RL, Quagliani D. Position of the American Dietetic Association: food and nutrition misinformation. J Am Diet Assoc 2002 Feb;102(2):260-266. [Medline: 11846124]

13. World Health Organization. 2008-2013 Action Plan for the Global Strategy for the Prevention and Control of Noncommunicable Diseases. Geneva: World Health Organization; 2008. URL: http://whqlibdoc.who.int/publications/2009/ 9789241597418_eng.pdf?ua=1 [accessed 2015-03-10] [WebCite Cache ID 6Wugdfxn0]

14. World Health Organization. Global Action Plan for the Prevention and Control of Noncommunicable Diseases 2013-2020. Geneva: World Health Organization; 2013. URL: http://apps.who.int/iris/bitstream/10665/94384/1/9789241506236_eng. pdf?ua=1 [accessed 2015-03-10] [WebCite Cache ID 6Wuh0bRHT]

15. National Health and Medical Research Council. Eat for Health: Australian Dietary Guidelines. Canberra: Commonwealth of Australia; 2013. URL: http://www.eatforhealth.gov.au/sites/default/files/files/the_guidelines/ n55_australian_dietary_guidelines.pdf [accessed 2015-03-09] [WebCite Cache ID 6WtXXmGt4]

16. National Health and Medical Research Council. Dietary Guidelines for Australian Adults. Canberra: Commonwealth of Australia; 2003 Apr 10. URL: https://www.nhmrc.gov.au/_files_nhmrc/publications/attachments/n33.pdf [accessed 2015-03-09] [WebCite Cache ID 6WtYEZGaH]

17. National Health and Medical Research Council. Dietary Guidelines for Children and Adolescents in Australia Incorporating the Infant Feeding Guidelines for Health Workers. Canberra: Commonwealth of Australia; 2003 Apr 10. URL: https://www. nhmrc.gov.au/_files_nhmrc/publications/attachments/n34.pdf [accessed 2015-03-09] [WebCite Cache ID 6WtYZPQoN]

18. National Health and Medical Research Council. Dietary Guidelines for Older Australians. Canberra: Commonwealth of Australia; 1999 Nov. URL: http://www.nhmrc.gov.au/_files_nhmrc/file/publications/synopses/withdrawn/n23.pdf [accessed 2015-03-09] [WebCite Cache ID 6WtXxHSsY]

19. Department of Health, Western Australia. Western Australian Health Promotion Strategic Framework 2012-2016. Perth: Chronic Disease Prevention Directorate, Department of Health Western Australia; 2012. URL: http://www. public.health.wa.gov.au/cproot/4462/2/wa_health_promotion_strategic_framework.pdf [accessed 2015-03-09] [WebCite Cache ID 6WtZDpNDZ]

20. Cugelman B, Thelwall M, Dawes P. Online interventions for social marketing health behavior change campaigns: a meta-analysis of psychological architectures and adherence factors. J Med Internet Res 2011;13(1):e17 [FREE Full text] [doi: 10.2196/jmir.1367] [Medline: 21320854]

21. Bennett GG, Glasgow RE. The delivery of public health interventions via the Internet: actualizing their potential. Annu Rev Public Health 2009;30:273-292. [doi: 10.1146/annurev.publhealth.031308.100235] [Medline: 19296777]

22. 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.

23. Chou WS, Prestin A, Lyons C, Wen K. Web 2.0 for health promotion: reviewing the current evidence. Am J Public Health 2013 Jan;103(1):e9-18. [doi: 10.2105/AJPH.2012.301071] [Medline: 23153164]

24. Duggan M, Ellison N, Lampe C, Lenhart A, Madden M. Social Media Update. Washington, DC: Pew Internet & American Life Project; 2014 Jan 09. URL: http://www.pewinternet.org/files/2015/01/PI_SocialMediaUpdate20144.pdf [accessed 2015-03-10] [WebCite Cache ID 6Wuj6zrJC]

25. Elkin N. How America Searches: Health and Wellness. New York: iCrossing; 2008 Jan. URL: http://www.icrossing.com/ sites/default/files/how-america-searches-health-and-wellness.pdf [accessed 2015-03-10] [WebCite Cache ID 6WujQoBqN]

26. Thackeray R, Crookston BT, West JH. Correlates of health-related social media use among adults. J Med Internet Res 2013;15(1):e21 [FREE Full text] [doi: 10.2196/jmir.2297] [Medline: 23367505]

27. Watson R, Wyness L. 'Don't tell me what to eat!' - Ways to engage the population in positive behaviour change. Nutrition Bulletin 2013 Feb 07;38(1):23-29. [doi: 10.1111/nbu.12003]

28. Centers for Disease Control and Prevention. The Health Communicator's Social Media Toolkit. 2011 Jul. URL: http://www.cdc.gov/healthcommunication/ToolsTemplates/SocialMediaToolkit_BM.pdf [accessed 2015-03-10] [WebCite Cache ID 6WuoYDWxN]

29. Ledford CJW. Changing channels: a theory-based guide to selecting traditional, new, and social media in strategic social marketing. Social Marketing Quarterly 2012 Sep 28;18(3):175-186. [doi: 10.1177/1524500412460671]

30. Meng X, Daly A, Pollard CM, Binns CW. Community attitudes toward breastfeeding in public places among Western Australia Adults, 1995-2009. J Hum Lact 2013 May;29(2):183-189. [doi: 10.1177/0890334413478835] [Medline: 23504475]

31. Australian Bureau of Statistics. 2012. Estimated resident population by single year of age, Western Australia table 55 URL: http://www.abs.gov.au/AUSSTATS/abs@.nsf/DetailsPage/3101.0Dec%202011?OpenDocument [accessed 2015-03-16] [WebCite Cache ID 6X41s6QH8]

32. Australian Bureau of Statistics. 2008. 2039.0-Information Paper: An Introduction to Socio-Economic Indexes for Areas (SEIFA), 2006 URL: http://www.abs.gov.au/ausstats/abs@.nsf/mf/2039.0/ [accessed 2015-03-16] [WebCite Cache ID 6X41OZ8bZ]

33. Australian Bureau of Statistics. 2005. 8146.0-Household Use of Information Technology, Australia, 2004-05 URL: http://www.abs.gov.au/AUSSTATS/abs@.nsf/allprimarymainfeatures/D299544DB0CA7EF0CA2572440077F833 [accessed 2015-03-09] [WebCite Cache ID 6WtaRJ5Ya]

34. Pew Research Center. 2014. Internet use over time URL: http://www.pewinternet.org/data-trend/internet-use/internet-use-over-time/ [accessed 2015-03-10] [WebCite Cache ID 6WujkljbR]

35. Visschers VH, Hartmann C, Leins-Hess R, Dohle S, Siegrist M. A consumer segmentation of nutrition information use and its relation to food consumption behaviour. Food Policy 2013 Oct;42(4):71-80. [doi: 10.1016/j.foodpol.2013.07.003]

36. Nölke L, Mensing M, Krämer A, Hornberg C. Sociodemographic and health-(care-)related characteristics of online health information seekers: a cross-sectional German study. BMC Public Health 2015;15:31 [FREE Full text] [doi: 10.1186/s12889-015-1423-0] [Medline: 25631456]

37. Statistics Canada. 2013 Nov 26. Canadian Internet Use Survey, 2012 URL: http://www.statcan.gc.ca/daily-quotidien/131126/dq131126d-eng.htm [accessed 2015-07-07] [WebCite Cache ID 6Zq78IFh5]

38. Zickuhr K, Smith A. Home Broadband 2013. Washington, DC: Pew Internet & American Life Project; 2013 Aug 26. URL: http://www.pewinternet.org/2013/08/26/home-broadband-2013/ [accessed 2015-07-07] [WebCite Cache ID 6Zq7Qrxsw]

39. File T, Ryan C. Computer and Internet Use in the United States. American Community Survey Reports, ACS-28, US. Washington, DC: US Census Bureau; 2014. URL: http://www.census.gov/content/dam/Census/library/publications/2014/acs/acs-28.pdf [WebCite Cache ID 6Zq7l5CK9]

40. Statistics Canada. 2010. Canadian Internet use survey, Internet use by individuals, by type of activity (Internet users at home) URL: http://www.statcan.gc.ca/tables-tableaux/sum-som/l01/cst01/comm29a-eng.htm [accessed 2015-03-10] [WebCite Cache ID 6WukwpMy3]

41. Modave F, Shokar NK, Peñaranda E, Nguyen N. Analysis of the accuracy of weight loss information search engine results on the internet. Am J Public Health 2014 Oct;104(10):1971-1978. [doi: 10.2105/AJPH.2014.302070] [Medline: 25122030]

42. Beck F, Richard J, Nguyen-Thanh V, Montagni I, Parizot I, Renahy E. Use of the internet as a health information resource among French young adults: results from a nationally representative survey. J Med Internet Res 2014 May;16(5):e128 [FREE Full text] [doi: 10.2196/jmir.2934] [Medline: 24824164]

43. Fox S, Duggan M. Health Online 2013. Washington, DC: Pew Internet & American Life Project; 2013. URL: http://www.pewinternet.org/files/old-media//Files/Reports/PIP_HealthOnline.pdf [accessed 2015-03-10] [WebCite Cache ID 6WunncdT6]

44. Gallagher S, Doherty DT. Searching for health information online: characteristics of online health seekers. J Evid Based Med 2009 May;2(2):99-106. [doi: 10.1111/j.1756-5391.2009.01021.x] [Medline: 21348996]

45. Lewis S, Thomas SL, Blood RW, Castle D, Hyde J, Komesaroff PA. 'I'm searching for solutions': why are obese individuals turning to the Internet for help and support with 'being fat'? Health Expect 2011 Dec;14(4):339-350. [doi: 10.1111/j.1369-7625.2010.00644.x] [Medline: 21199200]

46. Puhl R, Heuer C. The stigma of obesity: a review and update. Obesity (Silver Spring) 2009 May;17(5):941-964. [doi: 10.1038/oby.2008.636] [Medline: 19165161]

47. Macdiarmid JI, Loe J, Kyle J, McNeill G. "It was an education in portion size". Experience of eating a healthy diet and barriers to long term dietary change. Appetite 2013 Dec;71:411-419. [doi: 10.1016/j.appet.2013.09.012] [Medline: 24076020]

48. Tobey LN, Manore MM. Social media and nutrition education: the food hero experience. J Nutr Educ Behav 2014;46(2):128-133. [doi: 10.1016/j.jneb.2013.09.013] [Medline: 24220043]

49. Case P, Cluskey M, Hino J. Journal of Extension. 2011. Online nutrition education: enhancing opportunities for limited-resource learners URL: http://www.joe.org/joe/2011december/rb5.php [accessed 2015-07-30] [WebCite Cache ID 6aQb3UUpU]

50. Ostry A, Young ML, Hughes M. The quality of nutritional information available on popular websites: a content analysis. Health Educ Res 2008 Aug;23(4):648-655 [FREE Full text] [doi: 10.1093/her/cym050] [Medline: 17897928]

51. Powell J, Inglis N, Ronnie J, Large S. The characteristics and motivations of online health information seekers: cross-sectional survey and qualitative interview study. J Med Internet Res 2011 Feb;13(1):e20 [FREE Full text] [doi: 10.2196/jmir.1600] [Medline: 21345783]

XSL•FO
RenderX

52. Horodyska K, Luszczynska A, van den Berg M, Hendriksen M, Roos G, De Bourdeaudhuij I, et al. Good practice characteristics of diet and physical activity interventions and policies: an umbrella review. BMC Public Health 2015;15:19 [FREE Full text] [doi: 10.1186/s12889-015-1354-9] [Medline: 25604454]

53. Korda H, Itani Z. Harnessing social media for health promotion and behavior change. Health Promot Pract 2013 Jan;14(1):15-23. [doi: 10.1177/1524839911405850] [Medline: 21558472]

54. Dooley JA, Jones SC, Iverson D. Using Web 2.0 for health promotion and social marketing efforts: lessons learned from Web 2.0 experts. Health Mark Q 2014;31(2):178-196. [doi: 10.1080/07359683.2014.907204] [Medline: 24878406]

55. Dane S, Mason C, O'Brien-McInally B. Household Internet use in Australia: A Study in Regional Communities. Report No. EP1310907. Australia: CSIRO; 2013.

56. Lefebvre CR. The new technology: the consumer as participant rather than target audience. Social Marketing Quarterly 2007 Aug 21;13(3):31-42. [doi: 10.1080/15245000701544325]

57. Australian Bureau of Statistics. Canberra: ABS; 2008. Australian Social Trends URL: http://www.abs.gov.au/AUSSTATS/abs@.nsf/Lookup/4102.0Chapter6102008 [accessed 2015-03-09] [WebCite Cache ID 6WupF1Vlf]

58. O'Mara B. Aged care, cultural and linguistic diversity and IT in Australia: a critical perspective. Intl J of Migration, H and SC 2014 Jun 10;10(2):73-87. [doi: 10.1108/IJMHSC-05-2013-0006]

59. O'Hara BJ, Phongsavan P, Venugopal K, Bauman AE. Characteristics of participants in Australia's Get Healthy telephone-based lifestyle information and coaching service: reaching disadvantaged communities and those most at need. Health Educ Res 2011 Dec;26(6):1097-1106 [FREE Full text] [doi: 10.1093/her/cyr091] [Medline: 21987479]

60. O'Hara BJ, Phongsavan P, Venugopal K, Eakin EG, Eggins D, Caterson H, et al. Effectiveness of Australia's Get Healthy Information and Coaching Service®: translational research with population wide impact. Prev Med 2012 Oct;55(4):292-298. [doi: 10.1016/j.ypmed.2012.07.022] [Medline: 22885323]

61. NSW Office of Preventive Health. The Second Year (2013-14) in Review. Liverpool, NSW: NSW Office of Preventive Health; 2014. URL: http://www.preventivehealth.net.au/uploads/2/3/5/3/23537344/143102_preventive_health_second_year_in_review_report_art.pdf [accessed 2015-03-09] [WebCite Cache ID 6WusgtfHS]

Abbreviations

BMI: body mass index
NMSS: Nutrition Monitoring Survey Series
SEIFA: Socio Economic Indexes For Areas
SNAP: Supplemental Nutrition Assistance Program

Original Paper

Automatically Detecting Failures in Natural Language Processing Tools for Online Community Text

Albert Park[1], PhD; Andrea L Hartzler[2], PhD; Jina Huh[3], PhD; David W McDonald[4], PhD; Wanda Pratt[5], PhD

[1]Department of Biomedical Informatics and Medical Education, School of Medicine, University of Washington, Seattle, WA, United States
[2]Group Health Research Institute, Seattle, WA, United States
[3]Department of Biomedical Informatics, University of California San Diego, San Diego, CA, United States
[4]Human Centered Design & Engineering, University of Washington, Seattle, WA, United States
[5]The Information School, University of Washington, Seattle, WA, United States

Corresponding Author:
Albert Park, PhD
Department of Biomedical Informatics and Medical Education
School of Medicine
University of Washington
Box SLU-BIME 358047, 850 Republican St, Building C
Seattle, WA, 98109-4714
United States
Phone: 1 206 743 7843
Fax: 1 206 616 3152
Email: alpark1216@gmail.com

Abstract

Background: The prevalence and value of patient-generated health text are increasing, but processing such text remains problematic. Although existing biomedical natural language processing (NLP) tools are appealing, most were developed to process clinician- or researcher-generated text, such as clinical notes or journal articles. In addition to being constructed for different types of text, other challenges of using existing NLP include constantly changing technologies, source vocabularies, and characteristics of text. These continuously evolving challenges warrant the need for applying low-cost systematic assessment. However, the primarily accepted evaluation method in NLP, manual annotation, requires tremendous effort and time.

Objective: The primary objective of this study is to explore an alternative approach—using low-cost, automated methods to detect failures (eg, incorrect boundaries, missed terms, mismapped concepts) when processing patient-generated text with existing biomedical NLP tools. We first characterize common failures that NLP tools can make in processing online community text. We then demonstrate the feasibility of our automated approach in detecting these common failures using one of the most popular biomedical NLP tools, MetaMap.

Methods: Using 9657 posts from an online cancer community, we explored our automated failure detection approach in two steps: (1) to characterize the failure types, we first manually reviewed MetaMap's commonly occurring failures, grouped the inaccurate mappings into failure types, and then identified causes of the failures through iterative rounds of manual review using open coding, and (2) to automatically detect these failure types, we then explored combinations of existing NLP techniques and dictionary-based matching for each failure cause. Finally, we manually evaluated the automatically detected failures.

Results: From our manual review, we characterized three types of failure: (1) boundary failures, (2) missed term failures, and (3) word ambiguity failures. Within these three failure types, we discovered 12 causes of inaccurate mappings of concepts. We used automated methods to detect almost half of 383,572 MetaMap's mappings as problematic. Word sense ambiguity failure was the most widely occurring, comprising 82.22% of failures. Boundary failure was the second most frequent, amounting to 15.90% of failures, while missed term failures were the least common, making up 1.88% of failures. The automated failure detection achieved precision, recall, accuracy, and F1 score of 83.00%, 92.57%, 88.17%, and 87.52%, respectively.

Conclusions: We illustrate the challenges of processing patient-generated online health community text and characterize failures of NLP tools on this patient-generated health text, demonstrating the feasibility of our low-cost approach to automatically detect those failures. Our approach shows the potential for scalable and effective solutions to automatically assess the constantly evolving NLP tools and source vocabularies to process patient-generated text.

XSL•FO
RenderX

Introduction

The Internet pervades our everyday life, including health care [1]. For instance, patients increasingly use the Internet for health information and peer support. In 2005, 80% of Internet users searched for health information online [2]. Just 6 years later, a quarter of Internet users living with a chronic condition sought information online from a peer with a similar condition [3]. Similarly, applications that make use of data generated by patients are increasing [4]. For example, micro-blogging (eg, Twitter) has been used to improve natural disaster and emergency response situations [5], and patient-generated data on the PatientsLikeMe website has been used to evaluate the effectiveness of a drug [6]. Moreover, patients have voiced great potential benefits of such patient-generated data with respect to their treatment decisions, symptom management, clinical management, and outcomes [7,8]. However, in these instances, the use of patient-generated data required manual analysis of textual data. Although these methods provide value, manual analysis does not scale to the growing size of patient-generated health data online. Moreover, for many research activities, the overwhelming amount of data remains a challenge.

One scalable approach to process text-based patient-generated data is natural language processing (NLP). An increasing number of researchers studying patient-generated text, such as in online health communities, have used statistical methods based on manually annotated datasets [9-15]. Utilizing statistical methods, researchers extracted cancer patient trajectories from patients' posts [9], estimated the level of social support in an online breast cancer community [10], predicted adverse drug reactions from health and wellness Yahoo! Groups [11], identified medically relevant terms [12], classified addiction phases [13], predicted individual at risks for depression [14], and discovered patient posts in need of expertise from moderators [15]. These methods can be highly effective in a given online community, but they either require tremendous upfront effort to manually annotate or do not provide semantic connections. Furthermore, maintenance and generalizability remain as major challenges for such statistical methods.

Existing biomedical NLP tools have the potential to be used immediately and promise to provide greater generalizability than statistical approaches while providing semantic connections. Researchers have developed various NLP techniques and applications in the biomedical domain. For example, the Clinical Text Analysis and Knowledge Extraction System (cTakes) [16] was developed to map concepts to medical ontologies from clinical notes. cTakes is specifically trained for clinical domains and consists of NLP components that can be executed in sequence. Also, the National Center for Biomedical Ontology [17,18] annotator identifies a term and maps it to ontological concepts from multiple knowledge resources to allow the use of integrated knowledge. Other applications have been developed primarily for specific uses, such as Medical

Language Extraction and Encoding System (MedLEE) [19], whose goal pertains to identifying specified conditions in radiology reports. However, MedLEE was later adapted as a decision support system for Columbia-Presbyterian Medical Center [20] and as a phenotypic information extractor (BioMedLEE) [21] from biomedical literature.

One of the most widely regarded NLP applications in biomedicine is MetaMap [22], which was developed by the National Library of Medicine (NLM). MetaMap uses computational linguistic techniques to identify words or phrases in text and map them to concepts in the NLM's Unified Medical Language System (UMLS). The UMLS is a collaborative effort to enable semantic interoperability among systems by connecting more than 1.3 million concepts from more than 100 biomedical vocabularies [23,24]. Three knowledge sources enable applications to utilize the UMLS: (1) the Metathesaurus, which connects synonymous concepts across vocabularies, (2) the Semantic Network, which is a hierarchical network of semantic types that are linked to every concept in the Metathesaurus, and (3) the SPECIALIST Lexicon, which provides the lexical information needed for NLP tools. Thus, MetaMap provides a semantic link between the words or phrases in text and a structured vocabulary that is used by many applications in biomedicine.

However, MetaMap and many other traditional biomedical NLP tools were developed to process biomedical literature and clinical notes, rather than patient-generated text in online communities. One of the biggest challenges in applying these biomedical NLP tools to a different type of text is the difference in vocabulary. For example, Zeng et al recognize differences in the vocabulary used by patients and clinicians [25]. Smith and Wicks manually evaluated patient-generated text from PatientsLikeMe and found that over 50% of patient-submitted symptoms did not map to the UMLS due to issues like misspellings and slang [26]. Although Keselman et al [27] reported fewer cases of unmapped terms from patient-generated online community posts than Smith and Wicks [26], the researchers recognize this remaining challenge as a significant problem.

Recognizing the differences in vocabulary, a number of efforts to expand the UMLS to include patient-generated text have been reported [25,27-31]. One of the biggest efforts is the open-access Collaborative Consumer Health Vocabulary Initiative (CHV) [25,27,31]. CHV is a collaborative effort to address differences in terminology by including layman-friendly terminology that is familiar to patients [25]. Although the terminology difference could theoretically be addressed by expanded vocabularies, it is questionable whether CHV can fully address other issues of patient-generated text, such as misspellings, community nomenclature, and Internet-oriented writing styles. To address this issue, Elhadad et al applied an unsupervised, semantics-based methods to detect community nomenclature including typical misspellings [32]. Although the

method is domain-independent, it accounts for only three semantic types.

The effort to process patient-generated text, such as email [29,30] and search queries [27,28], using biomedical NLP tools has also been reported. For example, Brennan and Aronson processed patient-authored emails using MetaMap and showed the potential of processing this patient-generated, informal text to identify UMLS concepts [29]. However, Brennan and Aronson identified only three types of errors: (1) overly granular parsing of phrases into separate terms (eg, splitting of the phrase "feeling nauseous"), (2) inappropriate mappings that are simply nonsensical or incorrect for the context (eg, a verb "back" being mapped to "body location or region back"), and (3) mismatches resulting from terms and semantic types having more than one meaning (eg, confusion between "spatial concept right" and "qualitative concept right") [29]. Zeng et al have mapped the UMLS concepts to patients' Internet search queries [28], and the study highlighted the difference between terminology structures in UMLS and mental model of patients.

These prior studies show that many have worked to improve biomedical NLP tools to process patient-generated text. As NLP technologies and source vocabularies continue to evolve, we need easy, low-cost methods to systematically assess the performance of those tools. Traditionally in NLP, evaluations involve a great deal of manual effort, such as creating a manually annotated dataset. Moreover, a new evaluation for different types of text requires additional annotated datasets, thus maintenance can often be difficult. Recognizing the potential benefits of performing a low-cost assessment of NLP tools, we explore automated methods to detect failures without producing annotated datasets. Given MetaMap's long history of use in biomedical contexts, its configurability, and its scalability, we apply our failure detection tool to MetaMap in processing patient-generated text from an online cancer community to demonstrate the feasibility of automatically detecting occurrence of failures. We first present the dataset and MetaMap configuration, followed by the specific methods and results for (1) characterizing failure types, (2) automated failure detection, and (3) manual performance evaluation of our automated failure detection approach.

Methods

Online Community Dataset and MetaMap Configuration

Our dataset consists of community posts from the CancerConnect website, an online cancer community for cancer patients, their families, friends, and caregivers to exchange support and advice. The dataset consists of a total of 2010 unique user members and 9657 user member–generated posts from March 2010 to January 2013.

We processed the online community posts with MetaMap version 2011AA and configured the word sense disambiguation feature and included only the top-ranked concept from the output. In the default setting, MetaMap suggests a number of candidate concepts with candidate scores indicating relationships among concepts found in the text. However, in real-world usage

on large amounts of text, considering multiple suggestions for each processed term could be overwhelming to assess manually. Thus, we assessed only the top-ranked scored concept to simulate how MetaMap would be used in real-world settings. However, we used default settings for all other options for generalizability. A single mapped term/concept served as the unit of analysis.

Methods for Characterizing Failures

To characterize the types of failures, we assessed MetaMap's output collaboratively through iterative rounds of manual review among the five authors. We reviewed the output following an open coding process [33] to identify emerging themes grounded in data. Because we did not know all possible failure types, we chose to use an inductive coding process, rather than a structured, reductive content analysis approach. In each iteration, we processed different patient-generated posts, and each author independently and manually evaluated the same sets of mapped terms by examining the corresponding UMLS concept definitions and semantic types. Then as a group, we reached a consensus through discussion when different verdicts were made. Based on the list of inaccurate mappings, we grouped each inaccurate mapping into failure types and went on to identify potential causes within each failure type through the open coding process. This second step addressed the gaps in previous literature by identifying a number of causes of the failure types and providing information needed to detect these failures automatically.

Results for Characterizing Failures

Overview

From our manual review, we characterized three types of failure: (1) boundary failures, (2) missed term failures, and (3) word ambiguity failures. A boundary failure occurred when a single coherent term was incorrectly parsed into multiple incomplete terms. A missed term failure occurred when a relevant term had not been identified. A word sense ambiguity failure occurred when a relevant term was mapped to a wrong concept. Within these three failure types, we discovered 12 causes of failures. In the sections below, we describe each type of failure and then identify potential causes within each failure type.

Boundary Failures

Boundary failures, in which a single coherent term is incorrectly parsed into multiple terms, are well documented in biomedical NLP literature [26,29,34-36]. In this literature, boundary failures are referred to as overly granular parsing [29] or split phrasing [34]. Our analysis expands our understanding with boundary failures associated with patient-generated text.

Our patient-generated text contained extensive descriptive phrases (eg, "feeling great") and colloquial language (eg, "chemo brain"), contrasting with typical biomedical text that usually contained concepts from standard terminologies. Theoretically, boundary failures can result from standard medical terminologies. However, descriptive phrases and colloquial language highlight the parsing problem of biomedical NLP because colloquial language and descriptive phrases that patients use in online health communities cannot all be included

in the UMLS. For instance, UMLS included "feeling sick" as a synonym of a concept, although a similar descriptive phrase "feeling great" was not included in the UMLS. Consequently in our analysis, "feeling sick" was recognized as one concept, while "feeling great" was parsed into two separate terms "Emotions" and "Large" delivering different interpretations than intended.

Boundary failures also occurred even when proper concepts were available in the UMLS. For instance, a colloquial term "chemo brain" was commonly used to describe the single concept of cognitive deterioration of cancer patients after chemotherapy. In our analysis, the term was recognized as two UMLS concepts—"chemotherapy" and "brain-body part"—even though UMLS contained a concept for "chemo brain". From our experience, we inferred that the lack of colloquial language and descriptive phrases concepts in the UMLS as well as standard medical terminologies parser were causing boundary failure when processing patient-generated text.

Missed Term Failures

Overview

Missed term failures occurred when a relevant term was not identified [26,34]. We extended the literature by identifying two causes of missed term failures associated with patient-generated text: (1) community-specific nomenclature and (2) misspellings.

Community-Specific Nomenclature

Community-specific nomenclature refers to members of a community using terms that either are commonly used in a different way elsewhere or not commonly used at all. In online communities, members frequently create their own nomenclature that, over time, can become vernacular that is well understood in the community [37]. Community nomenclature poses unique challenges and opportunities for NLP.

In particular, community nomenclature regularly referred to relevant health-related content but resulted in three major challenges. First, many of the community-specific terms were not found in the UMLS. For instance, "PC" referred to "Prostate Cancer"; however, this acronym was not contained in UMLS. Second, community nomenclature was typically context and community-specific. For instance, the acronym "BC" was used for "before cancer", "blood count", or "breast cancer" depending on the context. This type of ambiguous usage was also seen with commonly accepted abbreviations. For instance, "rad" was a common abbreviation for "radiation therapy" in the cancer community, but "rad" could also be used for "radiation absorbed dose", "reactive airway disease", "reactive attachment disorder", or "RRAD gene" depending on the community. Third, novel abbreviations and acronyms constantly showed up in our data, similar to what researchers of online communities found [37]. For instance, our dataset included newly emerged acronyms that were not included in the UMLS, such as "LLS" and "PALS" for "Leukemia and Lymphoma Society" and "Patient Advice and Liaison Service", respectively.

Misspellings

Previous research showed that patients made more medically related misspellings at a significantly higher rate compared to clinicians [25]. Misspellings in our dataset included typographical errors (eg, "docotor"), phonetic errors that could be associated with lack of familiarity with medical terms (eg, "byopsi" and "methastasis"), and colloquial language errors (eg, "hooooooot flash"). Biomedical NLP techniques were typically developed using the correct spelling in training models, thus relevant but misspelled terms were often unrecognized. These unrecognized terms comprised a type of missed exact match [34] that consequently become false negatives—terms that should have been recognized but were missed. Although previous research in health information query investigated methods to address misspellings of patient-generated medical terms [38,39], those methods had limitations because they required correctly spelled medical terms in the database and manual selection of terms among recommended terms.

Word Sense Ambiguity Failures

The most prevalent failure was word sense ambiguity, which occurred when a term was mapped to the wrong concept because the two concepts are spelled the same way, share the same acronym (eg, "apt", an acronym used for appointment was mapped to organic chemical "4-azido-7-phenylpyrazolo-(1,5a)-1,3,5-triazine"), or were spelled the same as one of their acronyms (eg, a verb "aids" was mapped to "Acquired Immunodeficiency Syndrome"). This failure had been identified in previous research [26,29,34-36], but these studies did not examine the causes of this failure. From our analysis, we identified nine causes of failure associated with processing patient-generated text: (1) abbreviations and contractions, (2) colloquial language, (3) numbers, (4) email addresses and Uniform Resource Locators (URL)s, (5) Internet slang and short message service (SMS) language, (6) names, (7) the narrative style pronoun "I", (8) mismapped verbs, and (9) inconsistent mappings (by word sense disambiguation feature). In the following sections, we describe each cause of word sense ambiguity failures in detail and identify associated semantic types where applicable.

Abbreviations and Contractions

Frequent use of standard abbreviations and contractions was common in our online health communities. Online community members frequently used contractions such as "I'd" or abbreviations such as "i.e." in their text. Although the use of these shortened forms was common in informal text, it could be a source of errors for many NLP tools. For example, MetaMap maps "I'd" to "Incision and drainage" and mapped "i.e." to "Internal-External Locus of Control Scale" due to partial matches with synonyms. Also, MetaMap was inconsistent with some of its correct mappings for abbreviations. For instance, abbreviations for some US states were mapped correctly (eg, "AK" and "WA"), whereas others were often missed even though they were in the UMLS (eg, "CA" and "FL") or were mismapped (eg, Virginia was mapped to "Alveolar gas volume" when written as "V.A.").

Colloquial Language

Colloquial language, such as "hi" was prevalent in our dataset and caused many failures. Although these terms are obvious to human readers, we found they were often mapped to incorrect terms in the UMLS. For instance, our previous example "hi", rather than being left unmapped, was mapped to "Hawaii", "ABCC8 gene", or "AKAP4 gene" because "hi" was a synonym for all three concepts. In our analysis, this failure was found with many semantic types; however, terms mapped to the semantic type of "Gene or Genome" were particularly troublesome because of their unusual naming conventions.

Numbers: Dates, Times, and Other Numbers Not Indicating Disease Status

Our online community posts often contained numbers that convey important information, such as a patient's disease status (eg, "stage 3 breast cancer"). Other times, numbers conveyed more logistical information, such as time of day and dates, which were misinterpreted. For instance, in the phrase, "I got there at 4:12pm", "12pm" was mapped to "Maxillary left first premolar mesial prosthesis" because it was a complete match for one of its synonyms in the UMLS. Numbers that were used to convey diagnostic information were crucial for the identity of many community members, and such information was often included in an automated signature line (eg, "stage 2 grade 3 triple negative breast cancer") at the end of posts. Numbers indicating dates and times often resulted in false positives, whereas health status numbers often resulted in a different failure type (ie, boundary failure caused by splitting a phrase). We saw this type of failure across many different semantic types, including "Amino Acid, Peptide, or Protein", "Finding", "Gene or Genome", "Intellectual Product", "Medical Device", "Quantitative Concept", and "Research Activity".

Email Addresses and URLs

Online community members frequently mentioned URLs and email addresses in our dataset. They often pointed to websites that they found useful and gave out email addresses to start private conversations. Parts of email addresses and URLs were incorrectly mapped in our analysis. For instance, "net" at the end of an email address was often mapped to the "SPINK5 gene" because one of its synonyms was "nets". Also, "en", a language code that referred the English language in URLs, incorrectly mapped to "NT5E gene" because one of its synonyms was "eN".

Internet Slang and SMS Language

Internet slang and SMS language, such as "LOL" (ie, "laugh out loud" or "lots of love") or "XOXO" (ie, hugs and kisses) are highly prevalent in online community text but not in typical biomedical texts. Although these terms should be obvious to human readers, our analysis showed that Internet slang and SMS language were often mapped to incorrect biomedical terms in the UMLS. In particular, Internet slang and SMS language were often mistaken for gene names, such as the mapping of "LOL" to the LOX1 gene and "XO" to the XDH gene. To manage the different variations of concepts, the UMLS included many synonyms of terms. Varieties of these synonyms overlapped with commonly used Internet slang and SMS language resulting in word sense ambiguity failure.

Names: First, Last, and Community Handles

The use of names is also prevalent in online community posts, particularly when posts address specific individuals. Community members also often include their first names in a signature line and call out other members by first names or community handles. In our analysis, common first names were often mistaken for UMLS concepts, such as "Meg" being mistaken for "megestrol", "Rebecca" for "becatecarin", "Don" for "Diazooxonorleucine", and "Candy" for "candy dosage form". Each individual name was a complete match for one of the UMLS concepts. We identified these mismatches across multiple semantic types, including "Antibiotic", "Biomedical or Dental Material", "Clinical Attribute", "Diagnostic Procedure", "Disease or Syndrome", "Finding", "Hormone", "Injury or Poisoning", "Laboratory Procedure", "Mental Process", "Pathologic Function", "Pharmacologic Substance", and "Sign of Symptom".

Narrative Style of Pronoun "I"

Patients share a wide variety of personal experiences in narrative form in online health communities. Thus, the use of the pronoun "I" is prevalent in community posts but is a source of misinterpretation. For example, over the course of the study we discovered that "I" is typically mapped to either "Blood group antibody I" or "Iodides", which belong to "Amino Acid, Peptide, or Protein", "Immunologic Factor", or "Inorganic Chemical" semantic types.

Mismapped Verbs

One of the most fundamental components of NLP tools is a part-of-speech (POS) tagger, which marks up words with their corresponding POS (eg, verb, noun, preposition) in a phrase, sentence, or paragraph. POS taggers are commonly used in NLP and have many different applications, such as phrase parsers. In our analysis, we discovered that MetaMap uses a POS tagger called MedPost SKR (Semantic Knowledge Representation) [40] to split text into phrases. However, it did not use the resulting POS information when mapping to the UMLS. Such POS failures could have been overlooked in previous studies using biomedical text due to the fact that words like "said" or "saw" were less prevalent in biomedical literature or even in clinical notes. For our online community dataset, MetaMap improperly mapped terms without discriminating between verbs and nouns. For instance, simple verbs used in past tense, like "said" and "saw", were mapped as the acronym, "SAID" (ie, Simian Acquired Immunodeficiency Syndrome) and "saw" (ie, a medical device). Verbs in the present tense were also problematic. For instance, "bow" and "snap" were mapped to "Genu varum" and "Snap brand of resin", respectively. We observed this type of failure across different semantic types, including semantic types where verbs were unexpected, such as "Antibiotic", "Biomedical or Dental Material", and "Pharmacologic Substance".

Inconsistent Mappings

Two great strengths of the UMLS are its broad coverage of concepts and its capacity to distinguish among concepts in fine detail. This ability to provide the precise meaning of concepts is valuable for many applications. However, this feature also became a source for inconsistent mappings despite similar usage

of terms in our analysis. For instance, the term "stage" was mapped to multiple concepts in our dataset. Community members often used the term "stage" to describe their cancer status (eg, "stage 4 ER+ breast cancer"). Despite the seemingly similar sentence structures and usage of the term in the sentence, our findings showed that MetaMap inconsistently mapped

"stage" to different UMLS concepts. Six different semantic types were identified for the UMLS concepts mapping to "stage" (Table 1). This is a known failure of MetaMap [34]; however, the severity of the failures shows that addressing word sense disambiguation in patient-generated text may require particular attention.

Table 1. Word sense ambiguity failures: inconsistent mappings of stage by MetaMap.

Sample sentence	Mapped term	UMLS concept	Concept unique identifiers	UMLS semantic type
"My father was diagnosed with stage 2b pancreatic cancer"	stage 2b	Stage 2B	C0441769	Classification
"I'm stage 4 SLL and stage 2 CLL"	stage	Tumor stage	C1300072	Clinical attribute
"I was dx last year at age 46 with Stage 1"	Stage 1	Stage level 1	C0441766	Intellectual product
"Almost seven years ago I was diagnosed with stage 1 breast cancer at age 36 ½"	Stage breast cancer	malignant neoplasm of breast staging	C2216702	Neoplastic process
"My friend was just diagnosed with Stage IV cancer"	stage	Stage	C1306673	Qualitative concept
"My mom was diagnosed 11/07 with stage IV inoperable EC"	stage	Phase	C0205390	Temporal concept

Methods for Automated Failure Detection

Overview

To explore automated methods for detecting the three types of failures we identified, we created a tool that applies combinations of dictionary-based matching [41-43] and NLP techniques [44-47]. We describe this detailed automatic detection process in the following sections.

Detecting Boundary Failures

Our tool detected failures caused by incorrectly splitting a phrase through a comparison of MetaMap's MedPost SKR parser [40], a biomedical text parser, and the Stanford Parser [45] (a general-purpose parser). First, we collected all adjacent terms that MetaMap mapped but MedPost SKR had parsed separately. Second, we used the Stanford Parser to determine whether adjacently mapped terms were part of the same phrase. If adjacently mapped terms were part of the same phrase, the combined term could deliver a more precise meaning, while individually they often deliver different meanings [29,34]. We

found this especially problematic if the combined term was a valid UMLS concept. For instance, we would collect "chemo brain" as a boundary failure caused by splitting a phrase. "Chemo" and "brain" were terms that appeared adjacent to one another in a sentence, and their combination—"chemo brain"—was a valid UMLS concept, but MetaMap split it into two separate terms. However, we also collected combined terms that were not in the UMLS because they were also cases of improperly splitting terms. Furthermore, the missing combined terms could provide valuable insight to completeness of the UMLS. For instance, both "double mastectomy" and "chemo curls" are important concepts that are frequently used by patients; however, these concepts are missing from the UMLS as shown in Table 2. The aforementioned steps to compare MetaMap's MedPost SKR parser with the Stanford Parser can detect these important but missing terms. In our detection, we used the shortest possible phrase identified by the Stanford Parser. The Stanford Parser parsed phrases as structure trees to indicate grammatical relations. In the structure tree, a shorter phrase was often part of a longer phrase and delivered more coherent meanings compared to a longer phrase.

Table 2. Examples of splitting a phrase failure.

Sample sentence	Ideally mapped UMLS concept	First mapped term (UMLS concept name)	Second mapped term (UMLS concept name)
"My mom had unknown primary and it was a PET scan that helped them find the primary."	PET/CT scan	PET (Pet Animal)	Scan (Radionuclide Imaging)
"It was removed and I have had stereotactic treatment along with 6 rounds of Taxol/Carbo completed in January 2012." [sic]	Stereotactic Radiation Treatment	Stereotactic (Stereotactic)	Treatment (Therapeutic Aspects)
"Had 25 internal rad treatments (along with cisplatin on day 1 and 25)." [sic]	Therapeutic Radiology Procedure	Rad (Radiation Absorbed Dose)	Treatments (Therapeutic Procedure)
"I am Triple Negative BC and there are no follow-up treatments for us TN's."	Triple Negative Breast Neoplasms	Triple (Triplicate)	Negative (Negative)
"My doc thinks I will probably end up having a double mastectomy"	None available	Double (Double Value Type)	Mastectomy (Mastectomy)
"I thought after 9 months my hair would be back but I have grown some type of hair that I am told is 'chemo curls'."	None available	Chemo (Chemotherapy Regimen)	Curls (Early Endosome)

Detecting Missed Term Failures

Overview

We identified two causes of missed term failures associated with processing patient-generated text. The following sections describe automatic detection of missed terms, specifically due to community-specific nomenclature and misspellings.

Community-Specific Nomenclature

Our tool detected missed terms due to abbreviations and acronyms in four steps. First, it ran MetaMap on the original text and then counted the total number of mappings. Second, it extracted common abbreviations and acronyms and their definitions using a simple rule-based algorithm [46], but where we manually verified the extracted terms. Third, it ran MetaMap again after replacing the extracted abbreviations and acronyms with their corresponding fully expanded terms. Finally, it calculated the difference in the total mappings between the original text and the updated text.

The simple rule-based algorithm by Schwartz and Hearst [46] has performed well in finding abbreviations and acronyms in documents [48,49]. We modified the algorithm to reflect typical writing styles of online community posts. The algorithm by Schwartz and Hearst uses (1) order of characters, (2) distance between abbreviations/acronyms and their definitions, and (3) presence of parentheses to find candidates for abbreviations/acronyms and their definitions. Our tool first identified completely capitalized words (with an exception of the last character due to pluralization) as candidate abbreviations/acronyms and then applied the algorithm to find its fully expanded form. Because online community members adopted community's abbreviations/acronyms, we saved this information and applied to other posts written in the same community even when the definition was not available. For instance, in the sentence, "My mother was diagnosed with Stage 3 Esophageal cancer (EC) earlier this year - EC also counts smoking and alcohol as two major aggravating factors and is an aggressive cancer", the poster defined EC once and then continued to use the acronym in place of esophageal cancer.

MetaMap could map esophageal cancer but not EC. Our tool used this algorithm to detect EC and its fully expanded form, esophageal cancer, then replaced EC with "esophageal cancer" to ensure the concept could be identified by MetaMap.

Misspellings

Our tool detected the prevalence of missed terms due to misspelling using three steps. First, it ran MetaMap on the original text and counted the total number of mappings. Second, it ran MetaMap again after correcting possible misspellings using Google's query suggestion service [50]. Finally, it calculated the difference in the mappings between the original text and the corrected text.

Detecting Word Sense Ambiguity Failure

We identified nine causes of word sense ambiguity failure associated with processing patient-generated text. In the following sections, we describe how to automatically detect the word sense ambiguity failures.

Abbreviations and Contractions

To detect word sense ambiguity failures due to abbreviations and contractions, we used an NLP tool called the Stanford POS Tagger [44], which assigns POS to terms in text. Our tool processed the data using the Stanford POS Tagger to count cases where a single mapped term was tagged with multiple POS. For instance, the Stanford POS Tagger would accurately tag "I'd" with two different POS, that is, the personal pronoun and modal.

Colloquial Language

Detecting word sense ambiguity failure caused by colloquial language is particularly challenging. We identified many of these failures by narrowing our focus to consider only the "gene or genome" semantic type because colloquial language failures were frequently mapped to this semantic type. Our tool automatically detected improperly mapped colloquial language by using an existing cancer gene dictionary—a list of genes known to be associated with cancer [43]—and counting the number of terms categorized as a "gene or genome" semantic type that were not in the cancer gene dictionary.

Numbers: Dates, Times, and Other Numbers Not Indicating Disease Status

To automatically detect improperly mapped dates and times, we implemented a number of rule-based regular expressions to detect times and dates that were not mapped as "Quantitative Concept" semantic type concepts. "Quantitative Concept" is the most appropriate semantic type based on how patients typically used numbers in our dataset. This resulted in counting the numbers mapped to "Amino Acid, Peptide, or Protein", "Finding", "Gene or Genome", "Intellectual Product", "Medical Device", and "Research Activity".

In our approach, we recognized two types of date or time expression that are problematic for MetaMap. The first type was a time expression containing the term "pm". The second type was a string of numbers that has been typically used to describe age, date, or time duration. For instance, "3/4" indicating March fourth was mapped to a concept describing distance vision: concept unique identifier (CUI) C0442757. We used specific regular expressions that focused on numbers with "am" or "pm", as well as a string of numbers with or without non-alphanumeric characters in between numbers to identify dates, times, and other numbers that do not indicate disease status.

Email Addresses and URLs

Our detection process for email addresses and URLs was completed using regular expressions to identify all the email addresses and URLs, and then we counted the number of terms that were mapped from email addresses or URLs. In our approach, we used specific regular expressions matching "@" and a typical structure of domain name (ie, a dot character followed by 2-6 alphabetic or dot characters) for identifying email addresses and "http" or a typical structure of domain name for identifying URLs.

Internet Slang and SMS Language

We detected improperly mapped Internet slang and SMS language using a 3-step process. First, we identified an Internet dictionary with a list of chat acronyms and text shorthand [41]. Second, we manually reviewed the list to remove terms that were also medical acronyms. In this process, we identified only three medical acronyms, "AML", "CMF", and "RX" and removed them from the list. Third, our tool automatically identified the terms in the text by matching them with the Internet slang/SMS language list.

Names: First, Last, and Community Handles

To identify improperly mapped names, we first combined a number of name dictionaries that consist of first names [42] with a list of community handles from our online community, CancerConnect. Then, our tool counted the number of mapped terms that matched one of the names in the combined list.

Narrative Style of Pronoun "I"

We identified a number of cases where the pronoun "I" was improperly assumed to be an abbreviation, such as for Iodine, because the NLP tool did not consider the contextual knowledge from the term's POS. One of the most fundamental components of NLP tools is a POS tagger, which marks up words with their corresponding POS (eg, verb, noun, preposition) in a phrase, sentence, or paragraph. "I" as an abbreviation for Iodine should be recognized as a noun, whereas "I" meaning the individual should be recognized as a pronoun by a POS tagger. Our tool used data derived from the Stanford POS Tagger [44] to count cases where the pronoun "I" was mapped to either the "Blood group antibody I" or "Iodides" concepts. We noticed that the pronoun "I" was sometimes tagged as a foreign word. We included those cases in our counts because it was a failure of the Stanford POS Tagger.

Mismapped Verbs

To identify the improperly mapped terms without discriminating between verbs and nouns, we used POS information from the Stanford POS Tagger [44] to count cases where a mapped verb term belonged to a semantic type that did not contain verbs. The 34 semantic types (eg, "Activity" and "Behavior") listed under the "Event" tree of the UMLS ontology could contain verbs; thus, we excluded verbs from these semantic types from our analysis. We considered all verbs in the "Entity" tree of the UMLS ontology as incorrect mappings. The "Entity" portion includes semantic types, such as "Biomedical or Dental Material", "Disease or Syndrome", "Gene or Genome", "Medical Device", "Pharmacologic Substance", for which we do not expect verbs. Thus, our tool detected cases where verbs were associated with the "Entity" tree of the UMLS ontology.

Inconsistent Mappings

Detecting word sense ambiguity failures leading up to this section consisted of cases where terms were consistently mapped improperly. However, for other word sense ambiguity failures, MetaMap inconsistently mapped terms, both correctly and incorrectly. The inconsistency was the result of poor performance by MetaMap's word sense disambiguation feature that was designed to select the best matching concepts out of many candidate concepts available in the UMLS. We detected inconsistent mappings by (1) assuming that patients used terms consistently, and (2) MetaMap accurately selecting the best matching concepts the majority of the time. For instance, in our online cancer community dataset, we assumed that patients always used the term "blood test" to convey the "Hematologic Tests" concept (CUI: C0018941), which was how MetaMap interpreted this term two thirds of the time, rather than the less frequent mapping to the "Blood test device" concept (CUI: C0994779). Based on these assumptions, we detected inconsistent mappings in two steps. First, we created a term frequency table based on a term's spelling and its CUI. Second, assuming the most frequently mapped CUI was the correct concept, we counted the number of cases where the term was mapped to less frequent CUIs.

Results for Automated Failure Detection

The automated methods detected that at least 49.12% (188,411/383,572) of MetaMap's mappings for our dataset were problematic. Word sense ambiguity failures were the most widely occurring, comprising 82.22% among the total detected failures. Boundary failures were the second most frequent, amounting to 15.90% among the total detected failures, while missed term failures were the least common, making up 1.88% of the detected failures. Table 3 summarizes the identified

failures as well as their causes and prevalence for automatic detection of MetaMap's failure on processing patient-generated text. Our process showed the feasibility of automated failure detection; hence showing the types of failures that our tool could identify in similar datasets processed with biomedical NLP tools.

We found that word sense ambiguity failures were not mutually exclusive, and several cases had multiple causes. Thus, in Table 3, the sum of percentages for individual failures exceeded 100%. For instance, an acronym "OMG" used for "Oh My God" was incorrectly mapped to "OMG gene". This particular failure was detected as both colloquial language as well as Internet slang

and SMS language failures. To avoid redundant counts, we detected 154,904 unique counts of word sense ambiguity failure, making up 82.22% of failures. In Table 3, we show both individual counts/percentages as well as the total unique counts/percentages to provide a precise overview of word sense ambiguity failures. Although these failures were recognized in prior studies on MetaMap [26,29,34-36], the studies had not presented automated methods for detecting these failures.

We manually evaluated the performance of our failure detection tool in two parts: overall performance evaluation and individual component level performance evaluation.

Table 3. Detecting MetaMap's failures on processing patient-generated text.

Failure type	Causes of failure	Count	Percentage of failure, %
1. Boundary failures	1.1 Splitting a phrase	29,965	15.90
2. Missed term failures	2.1 Community specific nomenclatures	1167	0.62
	2.2 Misspellings	2375	1.26
3. Word sense ambiguity failures	3.1 Abbreviations and contractions	416	0.22
	3.2 Colloquial language	4162	2.21
	3.3 Numbers	143	0.08
	3.4 Email addresses and URLs	1448	0.77
	3.5 Internet slang and SMS language	3442	1.83
	3.6 Names	10,061	5.34
	3.7 Narrative style of pronoun 'I'	61,119	32.44
	3.8 Mismapped verbs	51,193	27.17
	3.9 Inconsistent mappings	29,308	15.56
	Total number of unique word sense ambiguity failures	154,904	82.22
Total number of unique failures		188,411	

Methods for Performance Evaluation of Automated Failure Detection

We randomly selected 50 cases (ie, mappings) that our tool identified as incorrect mappings from each of the 12 causes of failures, totaling 600 cases that served as positive cases. We then randomly selected another 600 cases from the rest of the mappings not detected as incorrect mappings according to our tool to serve as the negative cases. We then mixed up the selected 1200 cases and manually assessed the accuracy of mappings through a blind procedure.

We also measured individual performance on each of the 12 detection techniques. We used the previously selected 600 negative cases and individual technique's 50 positive cases to assess the performance. For boundary failure, we examined whether the mapped terms could deliver precise conceptual meaning independent of additional phrases. For missed term failure, we investigated whether the tool had accurately corrected the spellings and verified the results of the new mappings. For word sense ambiguity failures, we examined whether MetaMap appropriately mapped terms based on the rest of the context. The unit of analysis was a single mapping, and we evaluated

our results using precision, recall, accuracy, and F1 score. Precision measures the proportion of predicted positive instances that are correct. Recall measures the proportion of positive instances that were predicted. Accuracy measures the percentages of correctly predicted instances among the total number of instances examined. F1 score is the weighted harmonic mean—reflecting both performance and balance—of precision and recall. In all measures, higher scores reflect better performance.

Results for Performance Evaluation

Overview

Table 4 shows the performance of the automatic failure detection tool. The failure detection tool achieved overall precision, recall, accuracy, and F1 score of 83.00%, 92.57%, 88.17%, and 87.52%, respectively. At the individual component level, methods using dictionary-based matching or regular expression matching performed more accurately than methods using existing NLP techniques. In the following sections, we discuss findings of individual component of the automatic failure detection tool and its performance.

Table 4. Performance (in %) of automatic failure detection and its individual component.

Failure type	Causes of failure	Precision	Recall	Accuracy	F1 score
1. Boundary failures	1.1 Splitting a phrase	82.00	78.85	96.78	80.39
2. Missed term failures	2.1 Community specific nomenclatures	88.00	100.00	99.02	93.62
	2.2 Misspellings	80.00	93.02	97.88	86.02
3. Word sense ambiguity failures	3.1 Abbreviations and contractions	82.00	95.35	98.20	88.17
	3.2 Colloquial language	100.00	100.00	100.00	100.00
	3.3 Numbers	100.00	100.00	100.00	100.00
	3.4 Email addresses and URLs	100.00	100.00	100.00	100.00
	3.5 Internet slang and SMS language	100.00	100.00	100.00	100.00
	3.6 Names	66.00	100.00	97.21	79.52
	3.7 Narrative style of pronoun "I"	100.00	100.00	100.00	100.00
	3.8 Mismapped verbs	32.00	100.00	94.43	48.48
	3.9 Inconsistent mappings	66.00	53.23	92.80	58.93
Total		83.00	92.57	88.17	87.52

Boundary Failure

Our automatic failure detection tool identified 15.90% of the total failures as due to splitting a phrase. The performance evaluation of this task achieved precision, recall, accuracy, and F1 score of 82.00%, 78.85%, 96.78%, and 80.39%, respectively. It is important to note that a single concept can produce multiple split phrase failures. For instance, the phrase "stage 4 Melanoma" was mapped to three concepts: "stage", "4", and "Melanoma". Two boundary failures occurred in this phrase. The first failure occurred between "stage" and "4"; the second failure occurred between "4" and "Melanoma". By focusing on a pair of mapped terms at a time, we correctly identified two failures that occurred in the phrase "stage 4 Melanoma". We considered only adjacent paired mappings because splitting a single coherent phrase into two or more UMLS concepts was clearly a more significant problem. However, split phrase failures could occur in non-paired mappings as well, and we are underestimating the prevalence of split phrases.

Detecting Community-Specific Nomenclature

Less than 1% of failures were due to community-specific nomenclature, and the automatic detection system achieved precision, recall, accuracy, and F1 score of 88.00%, 100.00%, 99.02%, and 93.62%, respectively. It should be noted that we underestimated the number of missed terms because the algorithm [46] can identify abbreviations or acronyms only if they were previously defined by members at some point. In addition, we would not recognize cases where MetaMap still missed the fully expanded term.

Detecting Misspellings

We automatically assessed that misspellings were responsible for 1.26% of failures. However, we observed few cases of incorrect assessment due to failures of Google's query suggestion service. For instance, some medications were incorrectly recommended. "Donesaub", a misspelling of "Denosumab" was mapped to "dinosaur". Furthermore, even

with correct recommendation, MetaMap did not always map to the right concept. For instance, "Wsihng" was correctly recommended to be "Wishing", but MetaMap mapped it to "NCKIPSD gene". Despite a few cases of incorrect assessment, the misspelling component performed relatively well, achieving precision, recall, accuracy, and F1 score of 80.00%, 93.02%, 97.88%, and 86.02%, respectively.

Detecting Abbreviations and Contractions

Improperly mapped abbreviations comprised less than 1% of failures. Although this was seldom, the automatic detection system performed relatively well, achieving precision, recall, accuracy, and F1 score of 82.00%, 95.35%, 98.20%, and 88.17%, respectively.

Detecting Colloquial Language

Incorrectly mapped "gene or genome" semantic types comprised 2.21% of failures, and the automatic detection system achieved precision, recall, accuracy, and F1 score of 100.00%, 100.00%, 100.00%, and 100.00%, respectively. With this process, we also detected terms like "lord" and "wish" that may not be perceived as colloquial language. Nevertheless, they were improperly mapped as "gene or genome" semantic type. It is also important to note that different disease-specific communities should utilize different gene dictionaries.

Detecting Numbers: Dates, Times, and Other Numbers Not Indicating Disease Status

Our automatic failure detection tool identified less than 1% of failures as improperly mapped numbers. The performance evaluation of this task achieved precision, recall, accuracy, and F1 score of 100.00%, 100.00%, 100.00%, and 100.00%, respectively. However, we are underestimating this failure prevalence because MetaMap improperly mapped more than half of the "Quantitative Concept" semantic type concepts in our dataset. We did not include this semantic type and underestimated this particular failure because few cases were correctly mapped.

Detecting Email Addresses and URLs

Improperly mapped email addresses or URLs comprised less than 1% of failures, and the automatic detection system achieved precision, recall, accuracy, and F1 score of 100.00%, 100.00%, 100.00%, and 100.00%, respectively. It is important to note that the basis for our manual assessments was how patients had intended to use the term. For instance, MetaMap mapped "org" at the end of a URL to "Professional Organization or Group" concept. Although assessment of such cases can be subjective, we followed the basic rule of reflecting patients' intentions.

Detecting Internet Slang and SMS Language

A total of 1.83% of failures resulted from Internet slang and SMS language terms. Like other dictionary-based matching techniques, our automatic detection system performed relatively well, accomplishing precision, recall, accuracy, and F1 score of 100.00%, 100.00%, 100.00%, and 100.00%, respectively.

Detecting Names: First, Last, and Community Handles

We automatically assessed that names accounted for 5.34% of failures. However, the name dictionary matching did not perform as well as other dictionary-based matching components. We discovered that unique but popular names, such as "Sunday", "Faith", and "Hope" were incorrectly mapped when used as nouns in a sentence. The name dictionary component achieved precision, recall, accuracy, and F1 score of 66.00%, 100.00%, 97.21%, and 79.52%, respectively.

Detecting Narrative Style of Pronoun "I"

We found that 32.44% of failures resulted from pronoun "I". Although the use of the pronoun "I" could be considered a part of colloquial language, we noted it as a different cause of failure due to its high frequency. The automatic detection system accomplished precision, recall, accuracy, and F1 score of 100.00%, 100.00%, 100.00%, and 100.00%, respectively.

Detecting Mismapped Verbs

We automatically assessed that mismapped verbs accounted for 27.17% of failures; however, the detecting mismapped verbs component performed poorly, achieving precision, recall, accuracy, and F1 score of 32.00%, 100.00%, 94.43%, and 48.48%, respectively. We discovered that although Stanford POS Tagger has identified verbs correctly, we made the false assumption that verbs did not belong to the entity part of the UMLS ontology. However, verbs like "lost" and "wait" belong to the "Functional Concept" semantic type, which is under the entity part of the UMLS tree. Thus, the detecting mismapped verbs component of our automatic failure detection tool incorrectly identified such verbs as failures.

Detecting Inconsistent Mappings

Our automatic failure detection tool identified 15.56% of the total failures due to inconsistent mappings. The performance evaluation of this task achieved precision, recall, accuracy, and F1 score of 66.00%, 53.23%, 92.80%, and 58.93%, respectively. We found two reasons for the relatively low precision. First, we did not account for cases where the most commonly mapped concept is not the correct mapping. For instance, in our dataset "radiation" was mapped to "radiotherapy research" (CUI:

C1524021) two-thirds of the time when community members actually meant "therapeutic radiology procedure" (CUI: C1522449). We incorrectly assessed if less frequent mappings were accurate. Second, we missed cases when correct mappings do not exist. For instance, the verb "go" was incorrectly but consistently mapped as "GORAB gene". In our automated failure detection analysis, our tool overlooked terms like "go" that were consistently mismapped.

Results

We characterized (1) boundary failures, (2) missed term failures, and (3) word ambiguity failures and discovered 12 causes for these failures in our manual review. We then used automated methods and detected that almost half of 383,572 MetaMap's mappings were failures. 82.22% of failures were word sense ambiguity. 15.90% of failures were boundary failure. 1.88% of failures were missed term failures. The automated failure detection achieved precision, recall, accuracy, and F1 score of 83.00%, 92.57%, 88.17%, and 87.52%, respectively.

Discussion

Principal Considerations

We first discuss challenges of using out-of-the-box biomedical NLP tools, such as MetaMap, to process patient-generated text. We then discuss the contributions and wider implications of our study for research activities that need to manage the constantly changing and overwhelming amount of patient-generated data. We end with summarizing our contributions to the medical Internet research community.

Figure 1 illustrates the challenges of processing patient-generated online health community text and the common failures of biomedical NLP tools on that text. In an example sentence, "Hi Meg, I wish my docotor would haven't said I'd have chemo brain. It's 12PM and I'm signing off! LOL Don", MetaMap produced 12 mappings, all of which were incorrect, and overlooked one misspelled term, "docotor", thus producing 13 failures.

Some of these failures are already known problematic failures of MetaMap [26,29,34-36]. Our findings extend prior work by identifying the causes for each failure type. Leveraging our understanding of those causes, we developed automated techniques that identified these previously highlighted failures effectively without having to produce manually annotated datasets. In demonstrating the feasibility of our automated failure detection tool, we delineated the use of easily accessible NLP techniques and dictionaries. These techniques can independently examine each failure type. We provided a detailed demonstration of our failure detection tool to allow researchers to select the parts of our approach that meet the focus of their NLP tool assessment. Additionally, our detection approach can be modified and used to rectify failures in NLP tools.

We focused our research on MetaMap; however, findings from our study can apply to other NLP tools in a similar manner. Few failure causes, such as inconsistency of word sense disambiguation feature, pertain more to MetaMap than other tools. However, any NLP tools that provide semantic

connections require a similar word sense ambiguity feature. Moreover, different NLP tools could excel in different areas, and our automated failure detection can cost-effectively highlight problematic areas. Similarly, our techniques for detecting failures could strengthen the performance of other NLP tools to process patient-generated text and more traditional types of text. For instance, the word sense ambiguity failure caused by neglecting POS information can also be problematic in different types of text, including biomedical literature. That failure might surface less frequently due to differences in sentence structure between the biomedical literature and patient-generated text. Nevertheless, it is a significant problem that applies to both types of text. Applying such POS information when mapping a term could increase the accuracy of the mappings from a variety of texts. Another example is the missed term failure caused by community nomenclature. MetaMap or other NLP tools will miss terms if particular synonyms are missing from the vocabulary source. Researchers could use the algorithm by Schwartz and Hearst [46] to collect various synonyms that are used in different domains and frequently update the vocabulary sources, such as UMLS. Furthermore, researchers could use the splitting-a-phrase detection technique to not only prevent boundary failures, but collect new medical jargon (eg, "chemo curls") and identify important concepts missing in the UMLS (eg, "double mastectomy").

The dictionary-based matching and NLP techniques used in our detection process were evaluated in previous studies [44-46]. However, these studies were conducted in different domains and have been shown to produce errors. Moreover, these tools were not evaluated for patient-generated text. In addition, the automated detection techniques are generally limited to the coverage of the UMLS and MetaMap's capability to map when accurate and full spellings were provided. To strengthen our findings, we evaluated each detection method as well as the overall performance (Table 4). However, our findings could be biased towards cancer community text and could be further strengthened by generalizing our results in different platforms or patient groups. It is also plausible that we have not encountered all failure types or causes for other patient-generated health data contexts.

Moreover, a number of updates were made for both the UMLS and MetaMap [51] since the beginning of our study. To maintain consistency, we continued to use the same versions of the UMLS and MetaMap. However, we used the latest version of MetaMap (2013) and the UMLS (2013AB) to process a sample of 39 posts that were illustrated here and then compared the results to our findings. Although some of these causes were amended in the new version of MetaMap, the majority (33/39, 85%) of the outcomes remained unchanged or changed but still problematic. All the improvements (6/39, 15%) were word sense ambiguity failures. The improved cases included (1) colloquial languages, (2) email addresses and URLs, (3) Internet slang and SMS language, (4) mismapped verbs, and (5) two cases of inconsistent mappings. Despite these improvements, none of the described 12 causes of failures had been completely addressed. The lack of significant improvement further illustrated the magnitude of the challenges of processing patient-generated text. Because technologies, source vocabularies, and characteristics of text continue to be updated in the field of NLP, the need for low-cost automated methods to assess the updates will continuously increase. We demonstrate the feasibility of such automated approaches in detecting common failures using MetaMap and patient-generated text.

Although our study focused on online health community text, the insights inform efforts to apply NLP tools to process various types of patient-generated text, including blogs or online journals, which share similar narrative writing styles and colloquial language. Moreover, Facebook and email provide conversational interactions similar to the interaction in online health communities. Tweets about emergency responses [5], public health trends [52], or clinical notes from electronic medical records (EMR) could contain a host of abbreviations that NLP tools could incorrectly map. Thus, our failure detection techniques could be applied in these other contexts to assess the capability of processing different types of patient-generated text.

Figure 1. Example failures that resulted from the application of MetaMap to process patient-generated text in an online health community (blue terms represent patient-generated text; black terms represent MetaMap's interpretation; and red terms represent failure type).
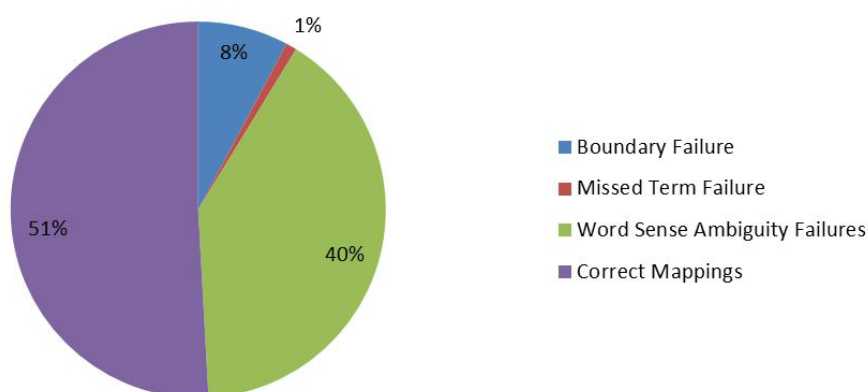
Conclusion

Processing patient-generated text provides unique opportunities. However, this process is fraught with challenges. We identified three types of failures that biomedical NLP tools could produce when processing patient-generated text from an online health community. We further identified causes for each failure type, which became the basis for applying automated failure detection methods using pre-validated NLP and dictionary-based techniques. Using these techniques, we showed the feasibility of identifying common failures in processing patient-generated health text, at a low cost. The value of our approach lies in helping researchers and developers quickly assess the capability of NLP tools for processing patient-generated text.

Conflicts of Interest

None declared.

References

1. Fox S, Rainie L. Pew Research Center Internet, Science & Tech. 2014 Feb 27. The Web at 25 in the US The overall verdict: The internet has been a plus for society and an especially good thing for individual users URL: http://www.pewinternet.org/2014/02/27/the-web-at-25-in-the-u-s/ [accessed 2015-04-23] [WebCite Cache ID 6Y5L8D2oE]

2. Fox S. Pew Research Center Internet, Science & Tech. 2005 May 17. Health Information online: Eight in ten internet users have looked for health information online, with increased interest in diet, fitness, drugs, health insurance, experimental treatments, and particular doctors and hospitals URL: http://www.pewinternet.org/2005/05/17/health-information-online/ [accessed 2015-04-23] [WebCite Cache ID 6Y5JQkmrG]

3. Fox S. Pew Research Center Internet, Science & Tech. 2011 Feb 28. Peer-to-peer healthcare: The internet gives patients and caregivers access not only to information, but also to each other URL: http://www.pewinternet.org/2011/02/28/peer-to-peer-health-care-2/ [accessed 2015-04-25] [WebCite Cache ID 6Y5JjmduI]

4. Eysenbach G. Medicine 2.0: social networking, collaboration, participation, apomediation, and openness. J Med Internet Res 2008;10(3):e22 [FREE Full text] [doi: 10.2196/jmir.1030] [Medline: 18725354]

5. Starbird K, Palen L. 'Voluntweeters': Self-Organizing by Digital Volunteers in Times of Crisis. : ACM; 2011 Presented at: ACM CHI Conference on Human Factors in Computing Systems; May 07-12, 2011; Vancouver, BC p. 1071-1080 URL: http://dl.acm.org/citation.cfm?id=1979102 [doi: 10.1145/1978942.1979102]

6. Wicks P, Vaughan TE, Massagli MP, Heywood J. Accelerated clinical discovery using self-reported patient data collected online and a patient-matching algorithm. Nat Biotechnol 2011 May;29(5):411-414. [doi: 10.1038/nbt.1837] [Medline: 21516084]

7. Wicks P, Massagli M, Frost J, Brownstein C, Okun S, Vaughan T, et al. Sharing health data for better outcomes on PatientsLikeMe. J Med Internet Res 2010;12(2):e19 [FREE Full text] [doi: 10.2196/jmir.1549] [Medline: 20542858]

8. Frost JH, Massagli MP. Social uses of personal health information within PatientsLikeMe, an online patient community: what can happen when patients have access to one another's data. J Med Internet Res 2008;10(3):e15 [FREE Full text] [doi: 10.2196/jmir.1053] [Medline: 18504244]

9. Wen M, Rose C. Understanding participant behavior trajectories in online health support groups using automatic extraction methods. : ACM; 2012 Presented at: 17th ACM international conference on Supporting group work; October 27–31, 2012; Sanibel Island, Florida p. 179-188 URL: http://dl.acm.org/citation.cfm?id=2389205 [doi: 10.1145/2389176.2389205]

10. Wang Y, Kraut R, Levine JM. To Stay or Leave ? The Relationship of Emotional and Informational Support to Commitment in Online Health Support Groups. : ACM; 2012 Presented at: ACM conference on Computer Supported Cooperative Work (CSCW); Feb 11-15, 2012; Seattle, Washington. [doi: 10.1145/2145204.2145329]

11. Chee BW, Berlin R, Schatz B. Predicting adverse drug events from personal health messages. AMIA Annu Symp Proc 2011;2011:217-226 [FREE Full text] [Medline: 22195073]

12. MacLean DL, Heer J. Identifying medical terms in patient-authored text: a crowdsourcing-based approach. J Am Med Inform Assoc 2013;20(6):1120-1127 [FREE Full text] [doi: 10.1136/amiajnl-2012-001110] [Medline: 23645553]

13. Maclean DL, Gupta S, Lembke A, Manning C, Heer J. Forum77: An Analysis of an Online Health Forum Dedicated to Addiction Recovery. : ACM; 2015 Presented at: 18th ACM conference on Computer supported cooperative work & social computing (CSCW); Mar 14-18, 2015; Vancouver, BC. [doi: 10.1145/2675133.2675146]

14. De Choudhury M, Gamon M, Counts S, Horvitz E. Predicting depression via social media. 2013 Presented at: 7th International AAAI Conference on Weblogs and SocialMedia (ICWSM); Jul 8-10, 2013; Boston, Massachusetts URL: http://www.aaai.org/ocs/index.php/ICWSM/ICWSM13/paper/viewFile/6124/6351

15. Huh J, Yetisgen-Yildiz M, Pratt W. Text classification for assisting moderators in online health communities. J Biomed Inform 2013 Dec;46(6):998-1005 [FREE Full text] [doi: 10.1016/j.jbi.2013.08.011] [Medline: 24025513]

XSL·FO

RenderX

16. Savova GK, Masanz JJ, Ogren PV, Zheng J, Sohn S, Kipper-Schuler KC, et al. Mayo clinical Text Analysis and Knowledge Extraction System (cTAKES): architecture, component evaluation and applications. J Am Med Inform Assoc 2010;17(5):507-513 [FREE Full text] [doi: 10.1136/jamia.2009.001560] [Medline: 20819853]

17. Musen MA, Noy NF, Shah NH, Whetzel PL, Chute CG, Story M, et al. The National Center for Biomedical Ontology. J Am Med Inform Assoc 2012;19(2):190-195 [FREE Full text] [doi: 10.1136/amiajnl-2011-000523] [Medline: 22081220]

18. Jonquet C, Shah NH, Musen MA. The open biomedical annotator. Summit on Translat Bioinforma 2009;2009:56-60. [Medline: 21347171]

19. Friedman C, Hripcsak G, DuMouchel W, Johnson SB, Clayton PD. Natural language processing in an operational clinical information system. Nat Lang Eng 2008 Sep 12;1(01):-. [doi: 10.1017/S1351324900000061]

20. Friedman C. Towards a comprehensive medical language processing system: methods and issues. Proc AMIA Annu Fall Symp 1997:595-599. [Medline: 9357695]

21. Chen L, Friedman C. Extracting phenotypic information from the literature via natural language processing. Stud Health Technol Inform 2004;107(Pt 2):758-762. [Medline: 15360914]

22. Aronson AR, Lang F. An overview of MetaMap: historical perspective and recent advances. J Am Med Inform Assoc 2010;17(3):229-236 [FREE Full text] [doi: 10.1136/jamia.2009.002733] [Medline: 20442139]

23. Humphreys BL, Lindberg DA, Schoolman HM, Barnett GO. The Unified Medical Language System: an informatics research collaboration. J Am Med Inform Assoc 1998;5(1):1-11 [FREE Full text] [Medline: 9452981]

24. Chen Y, Perl Y, Geller J, Cimino JJ. Analysis of a study of the users, uses, and future agenda of the UMLS. J Am Med Inform Assoc 2007;14(2):221-231 [FREE Full text] [doi: 10.1197/jamia.M2202] [Medline: 17213497]

25. Zeng Q, Kogan S, Ash N, Greenes RA. Patient and clinician vocabulary: how different are they? Stud Health Technol Inform 2001;84(Pt 1):399-403. [Medline: 11604772]

26. Smith CA, Wicks PJ. PatientsLikeMe: Consumer health vocabulary as a folksonomy. AMIA Annu Symp Proc 2008:682-686 [FREE Full text] [Medline: 18999004]

27. Keselman A, Smith CA, Divita G, Kim H, Browne AC, Leroy G, et al. Consumer health concepts that do not map to the UMLS: where do they fit? J Am Med Inform Assoc 2008;15(4):496-505 [FREE Full text] [doi: 10.1197/jamia.M2599] [Medline: 18436906]

28. Zeng QT, Crowell J, Plovnick RM, Kim E, Ngo L, Dibble E. Assisting consumer health information retrieval with query recommendations. J Am Med Inform Assoc 2006;13(1):80-90. [doi: 10.1197/jamia.M1820] [Medline: 16221944]

29. Brennan PF, Aronson AR. Towards linking patients and clinical information: detecting UMLS concepts in e-mail. J Biomed Inform 2003;36(4-5):334-341 [FREE Full text] [Medline: 14643729]

30. Smith CA, Stavri PZ, Chapman WW. In their own words? A terminological analysis of e-mail to a cancer information service. Proc AMIA Symp 2002:697-701 [FREE Full text] [Medline: 12463914]

31. Zeng QT, Tse T. Exploring and developing consumer health vocabularies. J Am Med Inform Assoc 2006;13(1):24-29 [FREE Full text] [doi: 10.1197/jamia.M1761] [Medline: 16221948]

32. Elhadad N, Zhang S, Driscoll P, Brody S. Characterizing the Sublanguage of Online Breast Cancer Forums for Medications, Symptoms, and Emotions. 2014 Presented at: AMIA Annual Fall Symposium; Nov. 15-19, 2014; Washington DC.

33. Strauss A, Corbin J. Basics of qualitative research. In: Basics of qualitative research: grounded theory procedures and techniques. Newbury Park, CA: Sage Publications; 1990.

34. Divita G, Tse T, Roth L. Failure analysis of MetaMap Transfer (MMTx). Stud Health Technol Inform 2004;107(Pt 2):763-767. [Medline: 15360915]

35. Pratt W, Yetisgen-Yildiz M. A study of biomedical concept identification: MetaMap vs. people. AMIA Annu Symp Proc 2003:529-533 [FREE Full text] [Medline: 14728229]

36. Kang N, Singh B, Afzal Z, van Mulligen EM, Kors JA. Using rule-based natural language processing to improve disease normalization in biomedical text. J Am Med Inform Assoc 2013;20(5):876-881 [FREE Full text] [doi: 10.1136/amiajnl-2012-001173] [Medline: 23043124]

37. Danescu-Niculescu-Mizil C, West R, Jurafsky D, Leskovec J, Potts C. No Country for Old Members: User Lifecycle and Linguistic Change in Online Communities. 2013 Presented at: 22nd Int Conf World Wide Web; May 13-17, 2013; Rio de Janeiro, Brazil p. 307-317.

38. McCray AT, Ide NC, Loane RR, Tse T. Strategies for supporting consumer health information seeking. Stud Health Technol Inform 2004;107(Pt 2):1152-1156. [Medline: 15360993]

39. McCray AT, Ide NC. Design and implementation of a national clinical trials registry. J Am Med Inform Assoc 2000;7(3):313-323 [FREE Full text] [Medline: 10833169]

40. Smith L, Rindflesch T, Wilbur WJ. MedPost: a part-of-speech tagger for bioMedical text. Bioinformatics 2004 Sep 22;20(14):2320-2321 [FREE Full text] [doi: 10.1093/bioinformatics/bth227] [Medline: 15073016]

41. Jansen E, James V. NetLingo: the Internet dictionary. 2002. URL: http://www.netlingo.com/ [accessed 2015-04-20] [WebCite Cache ID 6XwJyMzTG]

42. Bureau U census. Frequently Occurring Names in the US. 1990. URL: http://deron.meranda.us/data/census-derived-all-first.txt [accessed 2015-04-20] [WebCite Cache ID 6XwJ9Enlw]

XSL•FO

RenderX

43.   Bamford S, Dawson E, Forbes S, Clements J, Pettett R, Dogan A, et al. The COSMIC (Catalogue of Somatic Mutations in Cancer) database and website. Br J Cancer 2004 Jul 19;91(2):355-358 [FREE Full text] [doi: 10.1038/sj.bjc.6601894] [Medline: 15188009]

44.   Toutanova K, Klein D, Manning C, Singer Y. Feature-Rich Part-of-Speech Tagging with a Cyclic Dependency Network. 2003 Presented at: 2003 Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology - Volume 1; May 2003; Edmonton, AB p. 173-180 URL: http://dl.acm.org/citation.cfm?id=1073478 [doi: 10.3115/1073445.1073478]

45.   De Marneffe MC, MacCartney B, Manning CD. Generating Typed Dependency Parses from Phrase Structure Parses. 2006 Presented at: LREC; May 22-28, 2006; Genoa, Italy.

46.   Schwartz AS, Hearst MA. A simple algorithm for identifying abbreviation definitions in biomedical text. Pac Symp Biocomput 2003:451-462 [FREE Full text] [Medline: 12603049]

47.   Chapman WW, Bridewell W, Hanbury P, Cooper GF, Buchanan BG. A simple algorithm for identifying negated findings and diseases in discharge summaries. J Biomed Inform 2001 Oct;34(5):301-310 [FREE Full text] [doi: 10.1006/jbin.2001.1029] [Medline: 12123149]

48.   Baumgartner WA, Lu Z, Johnson HL, Caporaso JG, Paquette J, Lindemann A, et al. Concept recognition for extracting protein interaction relations from biomedical text. Genome Biol 2008;9 Suppl 2:S9 [FREE Full text] [doi: 10.1186/gb-2008-9-s2-s9] [Medline: 18834500]

49.   Jimeno-Yepes A, Berlanga-Llavori R, Rebholz-Schuhmann D. Ontology refinement for improved information retrieval. Information Processing & Management 2010 Jul;46(4):426-435. [doi: 10.1016/j.ipm.2009.05.008]

50.   Query Suggestion Service. URL: http://www.google.com/support/enterprise/static/gsa/docs/admin/70/gsa_doc_set/xml_reference/query_suggestion.html [accessed 2014-10-13] [WebCite Cache ID 6TIeuLgUB]

51.   MetaMap Updates. URL: http://metamap.nlm.nih.gov/MainDownload.shtml [accessed 2014-10-05] [WebCite Cache ID 6T6e318kp]

52.   Dredze M. How Social Media Will Change Public Health. IEEE Intell. Syst 2012 Jul;27(4):81-84 [FREE Full text] [doi: 10.1109/MIS.2012.76]

Abbreviations

CHV: Consumer Health Vocabulary
cTakes: Clinical Text Analysis and Knowledge Extraction System
CUI: concept unique identifiers
EC: esophageal cancer
EMR: electronic medical records
LOL: laugh out loud or lots of love
MedLEE: Medical Language Extraction and Encoding System
NLM: National Library of Medicine
NLP: natural language processing
POS: part of speech
Rad: radiation absorbed dose
SAID: Simian Acquired Immunodeficiency Syndrome
SKR: semantic knowledge representation
SMS: short message service
UMLS: Unified Medical Language System
URL: Uniform Resource Locator
XOXO: hugs and kisses

XSL•FO
RenderX

XSL•FO

RenderX

Original Paper

Use of the Blue Button Online Tool for Sharing Health Information: Qualitative Interviews With Patients and Providers

Dawn M Klein[1,2], MSW; Gemmae M Fix[3,4], PhD; Timothy P Hogan[3,5,6], PhD; Steven R Simon[7,8], MD; Kim M Nazi[9], PhD; Carolyn L Turvey[1,2], PhD

[1]Iowa City VA Health Care System, Comprehensive Access and Delivery Research and Evaluation (CADRE) Center, Iowa City, IA, United States

[2]University of Iowa, Carver College of Medicine, Department of Psychiatry, Iowa City, IA, United States

[3]Edith Nourse Rogers Memorial Veterans Hospital, Center for Healthcare Organization and Implementation Research (CHOIR), A VA HSR&D Center of Innovation, Bedford, MA, United States

[4]Boston University School of Public Health, Department of Health Policy and Management, Boston, MA, United States

[5]University of Massachusetts Medical School, Division of Health Informatics and Implementation Science, Department of Quantitative Health Sciences,, Worcester, MA, United States

[6]Edith Nourse Rogers Memorial Veterans Hospital, eHealth Quality Enhancement Research Initiative, National eHealth QUERI Coordinating Center, Bedford, MA, United States

[7]Veterans Affairs Boston Healthcare System, Section of General Internal Medicine, Boston, MA, United States

[8]Brigham and Women's Hospital, Division of General Medicine and Primary Care, Boston, MA, United States

[9]Veterans Health Administration, Veterans and Consumers Health Informatics Office, Office of Informatics & Analytics, Washington, DC, United States

Corresponding Author:
Dawn M Klein, MSW
Iowa City VA Health Care System
Comprehensive Access and Delivery Research and Evaluation (CADRE) Center
601 Highway 6 West
Iowa City, IA, 52246
United States
Phone: 1 3193535690
Fax: 1 3198874932
Email: dawn.klein@va.gov

Abstract

Background: Information sharing between providers is critical for care coordination, especially in health systems such as the United States Department of Veterans Affairs (VA), where many patients also receive care from other health care organizations. Patients can facilitate this sharing by using the Blue Button, an online tool that promotes patients' ability to view, print, and download their health records.

Objective: The aim of this study was to characterize (1) patients' use of Blue Button, an online information-sharing tool in VA's patient portal, My Health*e*Vet, (2) information-sharing practices between VA and non-VA providers, and (3) how providers and patients use a printed Blue Button report during a clinical visit.

Methods: Semistructured qualitative interviews were conducted with 34 VA patients, 10 VA providers, and 9 non-VA providers. Interviews focused on patients' use of Blue Button, information-sharing practices between VA and non-VA providers, and how patients and providers use a printed Blue Button report during a clinical visit. Qualitative themes were identified through iterative rounds of coding starting with an a priori schema based on technology adoption theory.

Results: Information sharing between VA and non-VA providers relied primarily on the patient. Patients most commonly used Blue Button to access and share VA laboratory results. Providers recognized the need for improved information sharing, valued the Blue Button printout, and expressed interest in a way to share information electronically across settings.

Conclusions: Consumer-oriented technologies such as Blue Button can facilitate patients sharing health information with providers in other health care systems; however, more education is needed to inform patients of this use to facilitate care coordination. Additional research is needed to explore how personal health record documents, such as Blue Button reports, can be easily shared and incorporated into the clinical workflow of providers.

XSL·FO
RenderX

Introduction

Care coordination can be defined as the deliberate organization of patient care activities among people involved in a patient's care to facilitate the appropriate delivery of health care services [1]. The Institute of Medicine identified care coordination as one of 20 priorities to improve the quality of health care in the United States [2]. The accurate and timely sharing of information is a critical component for effective coordination. In 2013, a survey conducted by the Office of the National Coordinator revealed that 1 in 3 individuals reported a gap in information exchange when seeking care for a medical problem. In addition, half of those who experienced a gap reported they had to tell the provider about their medical history because records were not received by another provider [3]. When information is lacking, patient safety may be compromised, duplicate services received (ie, unnecessary repeat laboratory testing), and health care costs increased [4-10].

One challenge to effective information sharing is that medical records are controlled and managed by health care organizations often resulting in a fragmented record for patients who receive care from multiple health care systems [11]. Electronic health records (EHRs) and health information exchanges (HIEs) are organizational health information technology solutions that are intended to help improve communication within and between care settings [12,13]. Although there has been was significant progress in EHR implementation, adoption of HIEs is variable across states and organizations [14]. In a study by Furukawa et al [15], only 14% of providers stated that they exchanged health information electronically with providers practicing outside of their health care system.

In contrast, electronic personal health records (PHRs) are managed by individuals [16,17]. Patient PHRs (patient portals) tethered to EHRs can help patients access information easily. The ability to do so has increased in recent years, in part due to meaningful use criteria that require health care systems to provide patients with access to their own health information [18-20]. Stage 2 of the meaningful use criteria further expands this mandate with requirements to provide patients the ability to transmit information securely [21]. Consumer-mediated exchange complements organizational HIE between care providers and systems. Given the requirements for consumer access to their health information, many organizations have adopted the Blue Button [22]. However, little is known about patient and provider attitudes about consumer-mediated exchange.

The Blue Button concept emerged in January 2010 at a Markle Consumer Engagement Workgroup with the goal of empowering consumers by providing them the ability to generate and download a single electronic file that contains their personal health information [23]. In August 2010, the Department of

Veterans Affairs (VA) launched their Blue Button in My Health*e*Vet, VA's combined patient portal and PHR [24]. Blue Button allows patients to easily view, print, and download their VA medical record data and self-entered information to create a report of their health information. Patients are able to customize the information they choose to include in their report by date range and data class. To date, more than 500 payers, providers, health-related associations, and others have taken the Blue Button Pledge to promote patient access to their own health data [25]. Although Blue Button awareness has grown, there is still a significant need for education because one-third of providers in a 2014 survey reported no familiarity with the Blue Button initiative [26]. With growing consumer access to their health information, it is important to examine patients and providers perceptions and experiences with Blue Button adoption and how it may be used to improve care coordination.

Recipients of VA health care benefits routinely receive care from both VA and non-VA providers. The 2011 Survey of Veteran Enrollees' Health and Reliance upon VA found that 77% of Veterans had alternative health insurance (ie, Medicare) [27]. In a study of rural Veterans, 75% indicated receiving care outside of the VA in the last year [28]. The 2014 Veterans Access, Choice, and Accountability Act [29] is also expected to increase care in civilian sectors making the need for effective care coordination even more critical for VA patients.

This study sought to understand how VA patients were adopting the Blue Button, information-sharing practices between VA and non-VA providers, and providers' thoughts about patients as the mediators of HIE. Further, with Blue Button and consumer-mediated sharing of health information as newer concepts, this study explored (1) if a report printed using the Blue Button could facilitate information sharing to support clinical care and (2) patient and provider preferences regarding receipt of this information.

Methods

Study Design

Qualitative interviews were conducted in 2012 as part of a larger evaluation of Blue Button adoption and use [30]. The focus of this study was on patient and provider experiences with PHRs and specifically the My Health*e*Vet Blue Button feature to identify (1) barriers and facilitators to adoption of Blue Button and (2) use of the Blue Button printout to improve coordination of care among all members in a Veteran's treatment team.

Participants

A rural Midwest and an urban Northeast VA health care system participated to represent both rural and urban locations. VA patients registered with a My Health*e*Vet account were identified and invited to participate by letter. For select sampling, interested participants were asked about prior Blue Button use

and if they received care from a non-VA provider. At the Midwest VA, non-VA providers were recruited through letter invitation from the local state association. Non-VA and VA providers were also identified with the assistance of project coinvestigators (CT, SS). Participants completed an informed consent process and all study procedures were approved by both Institutional Review Boards.

Using Rogers' Diffusion of Innovation theory [31] and the Unified Theory of Acceptance and Use of Technology (UTAUT) [32] as models, researchers formed interview questions (see Textbox 1 for examples of interview questions). These theories provided a flexible framework for determining key components influencing patients' adoption and use of Blue Button, and providers' perspectives toward use of Blue Button and PHRs for information sharing. Interview questions addressed how participants learned about Blue Button (knowledge/awareness), experiences using Blue Button (ease of use), and whether information printed from Blue Button was brought to health care visits (implementation), etc. Patients were asked about

communication between their providers and completed a baseline questionnaire about demographics and prior experience with My HealtheVet. Providers were asked about their experience coordinating care between organizations for VA patients, preferences for patients sharing information, and for input on the essential information to be included in a health summary. At the end of the interview, all participants were provided a sample 20-page Blue Button printout (Multimedia Appendix 1) based on a test patient with both VA EHR data (appointments, medication history, allergies, laboratory results, wellness reminders) and self-entered data (demographics, emergency contacts, health care providers, treatment facilities, health insurance, medications and supplements, allergies/adverse reactions, labs and tests, medical events, immunizations, vitals and readings, family health history, military health history). Participants critiqued the content and display and were asked for their opinions on using the document during a clinical visit. Interviews were conducted by trained research personnel primarily by phone and, when feasible, some provider interviews at the Midwest VA occurred in-person.

Textbox 1. Examples of semistructured interview questions for VA patients and providers.

VA Patients

- How do your doctors coordinate your medical care? How do they communicate?

- Tell me a little bit about your experience using My HealtheVet. What do you know about the Blue Button? Tell me about your experience using Blue Button.

- What do you think of the format of the information in the printout? Is it easy to understand?

- Have you ever brought information that you printed from Blue Button to a health care visit?

- If you were able to delegate your access to your personal health record / My HealtheVet account to your health care provider would you? (Meaning would you give your doctor the ability to log in to your My HealtheVet account?)

Providers

- Tell me about your experience giving and receiving information from [VA or non-VA] providers about patients' treatment (eg, medications).

- Are you familiar with personal health records / patient portals? Have you heard of the Blue Button?

- What is the essential information you want in a health summary? What do you think of the format of the Blue Button printout?

- Have you had experiences with patients who have brought in information from their personal health record, such as My HealtheVet?

- For patients who receive care from both VA and non-VA providers, what are your preferences on the best way for patients to share their health information with you?

- If a patient brought in a printout like this, what would you do with it?

- If a patient was able to delegate you access to their personal health record, do you think you would access it?

Analysis

Interviews were audio recorded, manually transcribed verbatim, and entered into the qualitative data software, NVivo 8 (QSR International Pty Ltd, Victoria, Australia) for coding. Codes were developed a priori based on concepts from Diffusion of Innovation [31] and UTUAT [32] theories. Additional, emergent themes were identified through grounded thematic approaches [33]. The research team (GF, CT, DK) developed codebooks specific to each participant group: non-VA providers, VA providers, and patients. For each codebook, the team systematically reviewed 2 to 3 interviews together and discussed key concepts. Codebooks were iteratively developed until no new codes were identified and saturation was reached. For the

provider interviews, fewer interviews were needed to reach saturation because their perspectives tended to be more similar whereas there was greater variability among patients. All interviews were then coded by a trained research assistant who worked closely with the team to consistently code the data. If there was clarification needed to address or revise codes, a team meeting was held to reach consensus.

Results

Participant Characteristics

A total of 34 patients completed interviews. See Table 1 for sample characteristics of participants. Of these, 24 VA patients reported using Blue Button previously and 22 received care

XSL·FO

RenderX

from a non-VA provider in addition to their VA provider. In all, 19 providers (9 non-VA and 10 VA) participated. Of the 10 VA providers interviewed, all were medical doctors (3 family practice, 4 internal medicine, 2 specialty care, 1 hospitalist). Of the 9 non-VA providers, 2 were nurses and 7 were medical doctors (3 family practice, 4 internal medicine/ primary care).

Table 1. Patient characteristics (N=34).

Sample characteristics	Participants
Age (years), mean (SD)	61.6 (10.2)
Gender (male), n (%)	33 (97)
Education, n (%)	
High school or some technical/college	20 (58)
College graduate or more	14 (42)
Income (US $; n=33), n (%)	
<$25,000	7 (21)
$25,000-$50,000	13 (39)
>$50,000	13 (39)
How long registered for My Health*e*Vet, n (%)	
≤1 year	13 (38)
2-3 years	15 (44)
≥4 years	6 (18)
How often use My Health*e*Vet, n (%)	
Less than once a month	11 (32)
About once a month	16 (47)
About once a week or more	7 (21)
Comfort using My Health*e*Vet	
Very comfortable	13 (38)
Somewhat comfortable	12 (35)
Neither comfortable/uncomfortable	5 (15)
Somewhat or very uncomfortable	4 (12)
Self-rated health, n (%)	
Excellent or very good	13(38)
Good	10 (29)
Fair or poor	11 (32)

Seven themes were identified: (1) knowledge of Blue Button; (2) ease of use, content, and readability of the Blue Button printout; (3) relative advantage of using Blue Button to access and share VA information; (4) perceived value of Blue Button; (5) patient experiences sharing VA health information; (6) provider perspectives on workflow and data quality; and (7) preferences sharing and receiving information. Table 2 (patient) and Table 3 (providers) summarize these themes with illustrative quotes. Of note, several themes for clinicians were similar regardless of organizational affiliation and these are indicated with "all" in Table 3.

Table 2. Summary of themes by patient responses with illustrative quotes.

Theme	Patient quote
Knowledge of Blue Button	
Found on their own exploring My Health*e*Vet	"...looking at My Health*e*Vet one day and then it just caught my eye...so I tried it."
Ease of use, content, and readability of Blue Button printout	
Mixed responses to the ease of use	"It's easy to use and it's self-explanatory."
	"Frustrating."
	"It's complicated."
Can be too long in length	"What I don't like is how much paper it takes up."
Relative advantage of Blue Button for accessing and sharing VA information	
Sharing information between providers relies on the patient	"...it's up to me to move the information back and forth."
Online access is easy and saves time	"It's easier than the phone and it's a timesaver as well."
	"It's better than me sending a release to [VA administrative office] and having them mail me...I could just go online [to get results] instead of having to wait for this giant document to come."
Perceived value of Blue Button	
Engaged in their own health care	"...the first time I used it I was really happy because I was participating in my health care. I mean you can actually see real time what's going on...which is really good. So it makes you part of the process."
Patient experience sharing VA health information	
Value in sharing information for time and money	"Saves a stick in the arm... saves them...the money and the time and effort...your lab results are just as good from 2 months ago, as they were from today to 3 weeks away."
Selective in what information is shared	"Typically just print the labs...here's the copy of my VA lab work..."
Preferences for sharing and receiving information	
Mixed response for preference to print or electronically share and preference based on what they perceive may be better for their provider	"I would prefer...[to] make the PDF from the Blue Button and then put it on my... iPad ...rather than printing off a bunch of paper."
	"Just bring in a copy...it's faster."
	"For [provider] I'd prefer to give her the paper copy because of her time...that way she can look it over when she's ready."
Value sharing self-reported information	"...with the over-the-counter medication and stuff like that they need to be aware of what was going on...the better treatment that I can get out of my provider is based on the more knowledge that they have, not out of an educated guess..."
Supportive of delegate access to VA providers; mixed support for delegate access to non-VA providers	"I'm confident in my providers and know they would maintain proper amount of security and are very ethical health care providers."
	"I don't know how many...non-VA providers you want snoopin' around in a VA record...I trust my doctor, but I don't know, it's a security type of thing with me."

Table 3. Summary of themes by provider responses with illustrative quotes.

Theme	Provider quote
Knowledge of Blue Button	
All: limited knowledge of Blue Button	"I've not seen that...this is not what I get from a VA patient. What I get from him is akin to an office visit summary." (non-VA)
Ease of use, content, and readability of Blue Button printout	
All: improve information display	"...it's not that the information is not useful, it's just that it's displayed over too many pages." (non-VA)
Non-VA: useful information; missing patient VA visit information (last visit note)	"You're not showing me the 'patient visit' here. The one thing that's missing." (non-VA)
Relative advantage of Blue Button for accessing and sharing VA information	
All: sharing information often relies on the patient	"I will print out stuff and give it to the patient, I say, 'Here, go give this to your urologist, okay?.'...and sometimes ...we tell the patients, 'You could have anything you want sent to whomever you want. Just go out to the business desk and those folks will take care of it.'" (VA)
Non-VA: difficult to get information from VA and patients sharing this information can help bridge the gap	"With VA, we get nothing...[W]e need something we have to call the VA or have the patient acquire it...Nothing is ever sent automatically from VA...and most of the time I don't even know that they see the VA..." (non-VA)
	"...we haven't had real good luck getting information from VA, so I think this is...better, the Blue Button." (non-VA)
Perceived value of Blue Button	
All: tool for patient education and value of self-report information	"Because people who are really reading or going through their records they are more involved in their health...they will learn more about their own health and their own medication." (VA)
	"Having the self-report is important, because that allows you to figure out what you think is going on and what they think is going on is different. And bringing together different sources, like looking at their pill bottles..." (VA)
Non-VA: improved efficacy	"...it would increase our accuracy and decrease our duplication of tests... it would make ...more economical sense for the patient insurance system as well." (non-VA)
Non-VA: abstract and incorporate relevant information in own electronic medical record	"It's not just this sort of scanned PDF, but rather something that becomes useable and actionable." (non-VA)
Provider perspectives on workflow and data quality	
All: mixed response to how it would impact workflow of clinical encounter, however information was valued	"It would help...I don't think it would add a lot of time...to have the information is important." (non-VA)
	"If they're very knowledgeable and could tell me all this verbally, then it probably doesn't necessarily save time...if this was a very long printout, it could take longer, but I don't think that's necessarily a bad thing. Because if you're getting a fuller picture...then I think that would be beneficial." (VA)
All: generally trust Blue Button report and self-entered information; however, may depend on the data reported and patient	"We ask patients to give us their history of what's happened to them, and we trust that. There are times we have to go corroborate that, so I see no reason why I wouldn't trust this any less than my encounter when I ask, 'So what's happened in the last year?'" (non-VA)
	"[the self-entered information] I get a little nervous...did they enter right? ...it's a data quality issue." (VA)
Preferences for sharing and receiving information	
All: prefer electronic receipt of data and easy sharing between systems	"What would be really ideal is if there were an interface between the community and the VA system where if a patient gets lab work done at the VA, or diagnostic studies done at the VA, or a colonoscopy done at the VA, right? Then, that stuff would come in and integrate with my system." (non-VA)

Theme	Provider quote
All: hard copy could be used for patient education	"I'd like them to bring it in their hand...We make notes on it, it goes back home with them. It's incredibly valuable for them to have stuff in their hands." (VA)
All: receipt of information in advance of clinical encounter	"It's always nice to have it ahead of the visit... then you can review it before the visit and highlight the important things that you want to address..." (non-VA)
All: mixed support for delegate access to patient's PHR	"In certain cases, I would...depends on the patient... and their problems." (VA)
	"...it spills on the wrong side of patients embracing responsibility for their health record. And therefore, it probably spills on the wrong side of where the liability is..." (non-VA)
All: open to patients logging in and sharing their personal health record data at a visit; however, possible work flow and technology barriers	"It's going to take some time, but chronic illness management requires some time. You need to take time and talk to people about these things during visits, and I think this would be a way of making it more efficient, not less...it's a tradeoff, you would save time by not doing it, but I think that not doing it is not a good option." (VA)
	"Right now, they can't log in to our computer, that would have to change...but, even so, I think that would sort of bog you down...in the office." (non-VA)

Knowledge of Blue Button

Most patients learned of the Blue Button simply seeing it on the My Health*e*Vet website. At the time of the study, VA and non-VA providers were generally unaware of the term "Blue Button" and had limited experience with patients using PHRs to share information with them. Among those who had heard of Blue Button, one non-VA provider did not know any details about it and a VA provider knew that patients could print a report using Blue Button, but had not actually received a printout from a patient.

Ease of Use, Content, and Readability of the Blue Button Printout

This theme focused on patients' ease of downloading a Blue Button file, content, and readability of the report. Patients primarily taught themselves how to use it with no specific training and there was variability with experiences in the ease of using the Blue Button. They liked the convenience of accessing their information online and that it was consolidated in one document, albeit the length of the printout was a concern for some.

Provider interviews focused on the sample Blue Button printout, critiquing the content and display of information. Overall, non-VA providers found the content useful; however, it was noted that some changes in formatting may improve the document. However, non-VA providers acknowledged that having the information outweighed the inconvenience of what they perceived as a cumbersome document due to its length.

When asked about the most important information to include in the Blue Button printout to inform clinicians, non-VA providers wanted a current medication list, laboratory test results, wellness reminders, immunizations, and allergies. In addition, there was interest in having the clinical note from the last visit available. VA providers reported much of the content redundant with information in the VA medical record and concurred with non-VA providers that the presentation of information could be improved.

Relative Advantage of Using Blue Button to Access and Share Information

For patients, this theme emerged in their discussion of the advantage of online access to their health information. This access could then facilitate information sharing. All interviewees (patients, VA providers, and non-VA providers) indicated that communication between providers primarily relies on the patient. Non-VA providers reported great difficulty getting health information from the VA and saw Blue Button as a possible solution to this problem. One non-VA provider, after viewing the sample printout, reported a preference for the Blue Button printout over other records from VA. VA providers already have access to much of the information in the Blue Button, so the relative advantage of Blue Button for information sharing was not evident. In contrast, non-VA providers who struggled to access VA information saw clear value.

Perceived Value of Blue Button

Two themes emerged under this domain: patient engagement and use of the Blue Button printout for clinical care and the health care system. With regards to patient engagement, patients liked having access to their health information. VA and non-VA providers reinforced this and expressed use of the printout as a tool that could help to identify gaps in understanding. Being better informed of all care can, in turn, help prevent errors or medication/test duplication and provide benefit from a cost perspective.

Non-VA providers were positive about the utility of the information for clinical decision making and indicated they would incorporate the information within their own medical records. One provider detailed the data sections (ie, allergies, medications) that he would integrate as structured data in their electronic record. In contrast to non-VA providers, VA providers saw little value in the printout for VA health information

because it would be redundant with information already accessible in VA's EHR. However, there was interest in reviewing the information that patient had self-entered into their PHR.

Patient Experiences Sharing Health Information From Blue Button

Patients reported mainly using the Blue Button for their own knowledge and most often reviewed laboratory results and appointments. There was limited experience sharing it with providers; however, for those who did share their Blue Button information with a non-VA provider, it was most often laboratory results. Those Veterans who have used the Blue Button for sharing information, tended to only print specific data rather than a long report. Most reported a favorable response sharing it, but one patient did have a negative experience that resulted in his perception that "nobody wants this...a doctor wants their own opinion."

Provider Perspectives on Workflow and Data Quality

The non-VA and VA providers interviewed had little to no experience with patients sharing a Blue Button printout; thus, to address the concept of using this document for clinical care, providers were asked for their perspectives on workflow and time burden if a patient were to present with a Blue Button printout at a visit. Results were mixed. Some expressed it would add time with others indicating it would be a time saver. Overall, providers expressed that having the information outweighed concerns about added workload and time.

There was also discussion if providers would "trust" a Blue Button document received from a patient and self-reported information in the report. Many providers indicated they would trust the accuracy of information accessed using Blue Button. In regards to patient self-reported information, some added the stipulation that similar to any self-report of information, it may depend on the patient and type of information reported. However, a few providers did note concern related to data quality for self-report.

Preferences for Sharing and Receiving Information

All patient and provider interviewees were also asked questions to explore their preferences for how information is shared to inform future guidance to VA patients using the Blue Button; this included receiving a printed hard copy either before or during a medical visit, electronic receipt, delegate access (assigning permission to allow the provider to sign into the patient's My Health*e*Vet account), or the patient logging in to their My Health*e*Vet account during an appointment.

Patients tended to want to share their information in a way that would be convenient for the provider. The majority were comfortable delegating My Health*e*Vet access to their VA providers, but some indicated it would be duplicative for VA information. Despite this, they saw value in sharing their self-entered data. For delegating access to non-VA providers, there was not consensus and patients expressed conflicting views. One patient who noted the value of self-entered data and stated he would delegate access to his VA provider reported that he would not delegate access to his non-VA provider due

to privacy concerns about sensitive health information. However, he would share a hard copy printout of selected information from his VA medical record if he deemed it necessary.

Patients who favored electronic information sharing (including logging in at an appointment to access their Blue Button) wanted to avoid unnecessary printing or misplacing of the document. Those who favored printing indicated that it was easier or faster.

Non-VA and VA providers expressed a desire for electronic receipt of information so that ideally it could be integrated into their EHR. Although it was often the preference, there was awareness that the exchange needed to be secure and a perception that this was not yet feasible with current technology. Some providers also wanted to receive the information in advance of an appointment. For others, hard copies shared during a visit were viewed as easier and one VA provider thought it could be used as a tool for patient education.

Provider support among both provider groups for delegate access was split. A non-VA provider agreed he would access the patient's My Health*e*Vet account if authorized and a VA provider also supported access depending on the patient situation. Alternatively, time and liability were significant concerns. Providers seemed more open to having a patient log in to their PHR during an appointment; however, workflow barriers included time and lack of computer access in exam rooms.

Discussion

In this sample of patients, VA providers, and non-VA providers, information sharing between providers relied primarily on the patient. Reports generated by the Blue Button feature of a PHR portal that contain both EHR data and patient self-entered information have great potential to facilitate care coordination in such contexts. Patients indicated some usability issues with the My Health*e*Vet Blue Button; however, they generally had favorable opinions of the technology. Providers recognized the need for improved information sharing. In particular, non-VA providers felt that having access to more VA health information would be of significant benefit. Many providers we spoke with expressed interest in a way to share information electronically across health care delivery settings. Although this study focused on VA patients receiving care from non-VA providers, meaningful use Stage 2 criteria promotes comparable consumer-mediated health information sharing between all health care settings.

These findings are consistent with other studies that found patients are responsible for sharing health information between providers [30,34-36] and limited use of the Blue Button printout despite its potential to improve information sharing [30]. The review of a sample Blue Button printout provided valuable insight about the potential impact of patients sharing their health information during a clinical encounter. Although the length of the report and time to review were potential barriers, in practice, patients who shared information using the Blue Button tended to be selective in choosing what they provided to their non-VA providers, consistent with other research [36,37]. Patients indicated they want to share information specifically relevant

to their care and in a way convenient to the provider. One concern is that patients may not know exactly what information to share. Although this may be perceived in some cases as patients intentionally omitting what might be clinically relevant information, it may in fact be a function of health literacy. Patient portals provide easier and timelier access to the EHR data in an unprecedented way. However, as noted earlier, because medical information has historically been managed at an institutional level, patients may have limited experience managing their own health information. This study and the prior literature suggest both a need to educate patients about using their patient portal to share information and for patient portals to be designed in ways that support patients' need to easily share critically important clinical information. In addition, it is equally important to inform providers about Blue Button and for providers to encourage patients to share their health data from providers in other care settings [38].

Non-VA provider experiences with care coordination are consistent with findings by Nayar et al [34]: 71% of non-VA providers surveyed reported they were rarely or never informed about VA care visits. Difficulty sharing information back to the VA was also apparent, with only 33% reporting that they had shared information with VA. HIEs are one mechanism that may help facilitate bidirectional exchange between systems [39], but adoption of HIEs is variable across states and organizations [14]. It is also critical for EHRs to support care coordination; however, EHRs often do not contain much information about outside care [40] and it is difficult to share across settings due to interoperability issues [12,41]. This supports the need for consumer-mediated information sharing because patients know where and when they will be receiving care. It also can empower patients by allowing them to choose what information to share and to be an engaged participant in their health care team.

This study provides support that providers and patients value and trust patient sharing of health record information and patient-generated data. Although patients can self-enter information in My HealtheVet, this information is not yet accessible to their VA providers. Data available from self-report in PHRs, remote monitoring devices, or personal wearables, can help inform providers' understanding of a patient's health between visits. Interest in these data is growing and it is important to develop meaningful ways for this data to be shared and integrated into workflow to help inform clinical care when relevant [42]. Efforts are underway for this information to be accessible to health care providers within the VA clinical information system and incorporated into clinical workflow.

Since these interviews, there have been continued enhancements to My HealtheVet, including increased patient access to information. In January 2013, clinical care notes were added to the Blue Button (VA OpenNotes) [43] and a Continuity of Care Document (CCD) / VA Health Summary became available. This document is a health summary that follows standards for interoperability with the goal of being integrated into other systems' EHRs. The VA CCD includes many of the essential data specified by non-VA providers; however, it does not yet provide the most recent clinic note, which was requested by non-VA providers in this study. The ability for patients to

securely transmit their CCD to an approved organization, non-VA provider, or application is currently in field testing.

Provider-focused HIE and consumer-mediated HIE can complement each other and provide benefit in different use cases. For example, in emergency department care, the immediate need for information may be served best by query or direct-based HIE models; whereas, in outpatient chronic care management, in which patients are often seen by multiple providers in different settings, consumer-mediated HIE can also effectively support care coordination. It is critical to educate patients about opportunities to participate in all types of HIE so they can make informed decisions about their preferences for sharing of their health information. As reflected by some participants in this study, some patients may not want to share all their health information with all providers. Expanding efforts in these areas supports the National Quality Strategy that focuses on patient engagement and effective communication as priorities to help meet the 3 broad aims of better, more affordable care for individuals and communities [44]. In a complex health care environment, multiple strategies are necessary to ensure information is available to provide safe and improved health care regardless of how the information is exchanged.

Limitations of this study include interviewing only patients who currently used My HealtheVet. Participants were asked to hypothesize how they would use the Blue Button printout for information sharing and clinical care, but it is unknown what they might do in actual practice. However, there is promising research reporting positive benefits of PHR use and patient engagement for health outcomes [45-49]. In addition, it was not required that patients have a non-VA provider; nevertheless, their use of the Blue Button to share self-reported information back to VA was relevant. The patients were predominantly older, white, male Veterans; younger or female patients may have different perspectives.

Although this study focused on the VA health care system, the results may be transferable to other settings where patients receive care from multiple health care systems and other organizations using Blue Button. In speaking to patients, VA providers, and non-VA providers, we gathered the perspectives from a range of stakeholders who are engaged in information sharing and use of tools such as Blue Button. Participants from geographically different regions of the country and different kinds of provider expertise (ie, family medicine, specialty care, and nursing) were included. In addition, meaningful diversity is also evident in the patient sample (ie, income, self-reported health status, and My HealtheVet experience). This variety of perspectives increases transferability [50] of findings across contexts.

More research is needed to examine whether patients sharing their health information with providers from different systems improves health care processes and outcomes. For example, does this consumer-mediated sharing improve medication reconciliation, reduce therapeutic duplication of medications, and/or reduce duplicate laboratory services or costs? In addition, future research should examine the impact on workflow for the provider receiving the information and, as transmit requirements from meaningful use 2 are implemented, examine how

information is received and incorporated into the EHR of the receiving provider or organization. As technology advances, effective processes must be developed in all care settings to enable all providers engaged in a patient's care to effectively share information for care coordination.

Acknowledgments

Authors' Contributions

CT was the principal investigator of this study and SS was coinvestigator. DK had primary responsibility for writing the manuscript. CT, DK, GF, SS, and KN contributed to the study design. CT, DK, GF, and SS were involved in study implementation and data collection. CT, DK, and GF were involved in data analysis and coding. All authors (DK, GF, TH, SS, KN, and CT) contributed to drafting of the manuscript and revising critically for intellectual content.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Sample Blue Button report.

[PDF File (Adobe PDF File), 289KB - jmir_v17i8e199_app1.pdf]

References

1. McDonald K, Sundaram V, Bravata D. Vol 7: Care Coordination. In: Shojania KG, McDonald KM, Wachter RM, editors. Closing the Quality Gap: A Critical Analysis of Quality Improvement Strategies. Technical Review 9. AHRQ Publication No 04(07)-0051-7. Rockville, MD: Agency for Healthcare Research and Quality; Jun 2007.

2. Adams K, Corrigan J. Priority Areas for National Action: Transforming Health Care Quality. Washington, DC: National Academies Press; 2003.

3. Patel V, Barker W, Siminerio E. ONC Data Brief No. 20: Individuals' Access and Use of their Online Medical Records Nationwide. Washington, DC: Office of the National Coordinator for Health Information Technology; 2014 Sep. URL: http://www.healthit.gov/sites/default/files/consumeraccessdatabrief_9_10_14.pdf [accessed 2015-04-29] [WebCite Cache ID 6YB0GfK4O]

4. Jha AK, Chan DC, Ridgway AB, Franz C, Bates DW. Improving safety and eliminating redundant tests: cutting costs in U.S. hospitals. Health Aff (Millwood) 2009;28(5):1475-1484 [FREE Full text] [doi: 10.1377/hlthaff.28.5.1475] [Medline: 19738266]

5. Kripalani S, LeFevre F, Phillips CO, Williams MV, Basaviah P, Baker DW. Deficits in communication and information transfer between hospital-based and primary care physicians: implications for patient safety and continuity of care. JAMA 2007 Feb 28;297(8):831-841. [doi: 10.1001/jama.297.8.831] [Medline: 17327525]

6. Stewart BA, Fernandes S, Rodriguez-Huertas E, Landzberg M. A preliminary look at duplicate testing associated with lack of electronic health record interoperability for transferred patients. J Am Med Inform Assoc 2010;17(3):341-344 [FREE Full text] [doi: 10.1136/jamia.2009.001750] [Medline: 20442154]

7. Trivedi AN, Grebla RC, Jiang L, Yoon J, Mor V, Kizer KW. Duplicate federal payments for dual enrollees in Medicare Advantage plans and the Veterans Affairs health care system. JAMA 2012 Jul 4;308(1):67-72. [doi: 10.1001/jama.2012.7115] [Medline: 22735360]

8. Jia H, Zheng Y, Reker DM, Cowper DC, Wu SS, Vogel WB, et al. Multiple system utilization and mortality for veterans with stroke. Stroke 2007 Feb;38(2):355-360 [FREE Full text] [doi: 10.1161/01.STR.0000254457.38901.fb] [Medline: 17194888]

9. Wolinsky FD, Miller TR, An H, Brezinski PR, Vaughn TE, Rosenthal GE. Dual use of Medicare and the Veterans Health Administration: are there adverse health outcomes? BMC Health Serv Res 2006;6:131 [FREE Full text] [doi: 10.1186/1472-6963-6-131] [Medline: 17029643]

10. Bodenheimer T. Coordinating care--a perilous journey through the health care system. N Engl J Med 2008 Mar 6;358(10):1064-1071. [doi: 10.1056/NEJMhpr0706165] [Medline: 18322289]

11. Elhauge E. The Fragmentation of US Health Care: Causes and Solutions. New York: Oxford University Press; 2010.

12. D'Amore JD, Mandel JC, Kreda DA, Swain A, Koromia GA, Sundareswaran S, et al. Are Meaningful Use Stage 2 certified EHRs ready for interoperability? Findings from the SMART C-CDA Collaborative. J Am Med Inform Assoc 2014;21(6):1060-1068 [FREE Full text] [doi: 10.1136/amiajnl-2014-002883] [Medline: 24970839]

13. Williams C, Mostashari F, Mertz K, Hogin E, Atwal P. From the Office of the National Coordinator: the strategy for advancing the exchange of health information. Health Aff (Millwood) 2012 Mar;31(3):527-536 [FREE Full text] [doi: 10.1377/hlthaff.2011.1314] [Medline: 22392663]

14. Adler-Milstein J, Jha AK. Health information exchange among US hospitals: who's in, who's out, and why? Healthcare 2014 Mar;2(1):26-32. [doi: 10.1016/j.hjdsi.2013.12.005]

15. Furukawa MF, King J, Patel V, Hsiao C, Adler-Milstein J, Jha AK. Despite substantial progress In EHR adoption, health information exchange and patient engagement remain low in office settings. Health Aff (Millwood) 2014 Sep;33(9):1672-1679. [doi: 10.1377/hlthaff.2014.0445] [Medline: 25104827]

16. Burrington-Brown J, Fishel J, Fox L, Friedman B, Giannangelo K, Jacobs E, AHIMA e-HIM Personal Health Record Work Group. Defining the personal health record. AHIMA releases definition, attributes of consumer health record. J AHIMA 2005 Jun;76(6):24-25. [Medline: 15986557]

17. Tang PC, Ash JS, Bates DW, Overhage JM, Sands DZ. Personal health records: definitions, benefits, and strategies for overcoming barriers to adoption. J Am Med Inform Assoc 2006;13(2):121-126 [FREE Full text] [doi: 10.1197/jamia.M2025] [Medline: 16357345]

18. Blumenthal D, Tavenner M. The "meaningful use" regulation for electronic health records. N Engl J Med 2010 Aug 5;363(6):501-504. [doi: 10.1056/NEJMp1006114] [Medline: 20647183]

19. Ahern DK, Woods SS, Lightowler MC, Finley SW, Houston TK. Promise of and potential for patient-facing technologies to enable meaningful use. Am J Prev Med 2011 May;40(5 Suppl 2):S162-S172. [doi: 10.1016/j.amepre.2011.01.005] [Medline: 21521591]

20. Kruse CS, Bolton K, Freriks G. The effect of patient portals on quality outcomes and its implications to meaningful use: a systematic review. J Med Internet Res 2015;17(2):e44 [FREE Full text] [doi: 10.2196/jmir.3171] [Medline: 25669240]

21. Office of the National Coordinator. 2014. ONC 2014 Edition EHR Certification Criteria Grid Mapped to Meaningful Use Stage 2 URL: http://www.healthit.gov/sites/default/files/2014editionehrcertificationcriteria_mustage2.pdf [accessed 2014-09-08] [WebCite Cache ID 6YAE4176Ve0]

22. Blue Button. 2014. URL: http://healthit.gov/patients-families/your-health-data [accessed 2015-06-30] [WebCite Cache ID 6ZfmHCl0E]

23. Nazi K. US Department of Veterans Affairs. 2010. VA's Blue Button: empowering people with their data URL: http://www.blogs.va.gov/VAntage/866/vas-blue-button-empowering-people-with-their-data/ [accessed 2015-06-24] [WebCite Cache ID 6ZYKbil1k]

24. Hogan TP, Nazi KM, Luger TM, Amante DJ, Smith BM, Barker A, et al. Technology-assisted patient access to clinical information: an evaluation framework for blue button. JMIR Res Protoc 2014;3(1):e18 [FREE Full text] [doi: 10.2196/resprot.3290] [Medline: 24675395]

25. HealthIT.gov. 2014. About the Blue Button Pledge URL: http://www.healthit.gov/patients-families/pledge-about [accessed 2015-06-25] [WebCite Cache ID 6ZYJ5cFOE]

26. Narcisi J. Workgroup for Electronic Data Interchange Blue Button Initiative Survey Results. 2015. URL: http://wedi.org/docs/public-policy/full-comment-letter-and-summary-of-survey-results.pdf?sfvrsn=0 [accessed 2015-06-25] [WebCite Cache ID 6ZYKLHSwW]

27. Department of Veterans Affairs. 2011 Survey of Veteran Enrollees' HealthReliance upon VA. 2012 Mar. URL: http://www.va.gov/healthpolicyplanning/soe2011/soe2011_report.pdf [accessed 2015-06-25] [WebCite Cache ID 6ZYKPbAS4]

28. Nayar P, Apenteng B, Yu F, Woodbridge P, Fetrick A. Rural veterans' perspectives of dual care. J Community Health 2013 Feb;38(1):70-77. [doi: 10.1007/s10900-012-9583-7] [Medline: 22772840]

29. Veterans Access, Choice, and Accountability Act of 2014. 2014. URL: http://www.gpo.gov/fdsys/pkg/BILLS-113hr3230enr/pdf/BILLS-113hr3230enr.pdf [accessed 2015-04-30] [WebCite Cache ID 6YAFJFqrj]

30. Turvey C, Klein D, Fix G, Hogan TP, Woods S, Simon SR, et al. Blue Button use by patients to access and share health record information using the Department of Veterans Affairs' online patient portal. J Am Med Inform Assoc 2014;21(4):657-663. [doi: 10.1136/amiajnl-2014-002723] [Medline: 24740865]

31. Rogers E. Diffusion of Innovations, 5th Edition. New York: Free Press; 2003.

32. Venkatesh V, Morris M, Davix G, Davis F. User acceptance of information technology: toward a unifed view. MIS Quarterly 2003;27(3):425-478.

33. Patton M. Qualitative Research and Evaluation Methods. Thousand Oaks, CA: Sage Publications; 2002.

34. Nayar P, Nguyen AT, Ojha D, Schmid KK, Apenteng B, Woodbridge P. Transitions in dual care for veterans: non-federal physician perspectives. J Community Health 2013 Apr;38(2):225-237. [doi: 10.1007/s10900-012-9604-6] [Medline: 22903804]

35. Howren MB, Cozad AJ, Kaboli PJ. Considering the issue of dual use in veterans affairs patients: implications & opportunities for improved communication & counseling. Health Commun 2015 Aug;30(8):838-842. [doi: 10.1080/10410236.2014.930299] [Medline: 25010429]

XSL•FO
RenderX

36.   Ancker JS, Witteman HO, Hafeez B, Provencher T, Van de Graaf M, Wei E. The invisible work of personal health information management among people with multiple chronic conditions: qualitative interview study among patients and providers. J Med Internet Res 2015;17(6):e137 [FREE Full text] [doi: 10.2196/jmir.4381] [Medline: 26043709]

37.   Rassin M, Zilcha L, Berger M, Silner D. Personal medical documents management--how patients perceive, keep and manage their medical documents: a qualitative study. Int J Nurs Stud 2007 Aug;44(6):862-868. [doi: 10.1016/j.ijnurstu.2006.03.005] [Medline: 16678180]

38.   Kerns JW, Krist AH, Longo DR, Kuzel AJ, Woolf SH. How patients want to engage with their personal health record: a qualitative study. BMJ Open 2013;3(7):e002931 [FREE Full text] [doi: 10.1136/bmjopen-2013-002931] [Medline: 23901027]

39.   Emani S, Yamin CK, Peters E, Karson AS, Lipsitz SR, Wald JS, et al. Patient perceptions of a personal health record: a test of the diffusion of innovation model. J Med Internet Res 2012;14(6):e150 [FREE Full text] [doi: 10.2196/jmir.2278] [Medline: 23128775]

40.   Institute for Health Technology Transformation. Population Health Management: A Roadmap for Provider-Based Automation in a New Era of Healthcare. New York: Institute for Health Technology Transformation; 2012. URL: http://ihealthtran.com/pdf/PHMReport.pdf [accessed 2015-04-30] [WebCite Cache ID 6YAFe1mfh]

41.   O'Malley AS, Grossman JM, Cohen GR, Kemper NM, Pham HH. Are electronic medical records helpful for care coordination? Experiences of physician practices. J Gen Intern Med 2010 Mar;25(3):177-185 [FREE Full text] [doi: 10.1007/s11606-009-1195-2] [Medline: 20033621]

42.   Deering M. Issue Brief: Patient-Generated Heatlh dataHealth IT. 2013. URL: http://www.healthit.gov/sites/default/files/pghd_brief_final122013.pdf [accessed 2015-06-25] [WebCite Cache ID 6ZYJkgYZ7]

43.   Nazi KM, Turvey CL, Klein DM, Hogan TP, Woods SS. VA OpenNotes: exploring the experiences of early patient adopters with access to clinical notes. J Am Med Inform Assoc 2015 Mar;22(2):380-389. [doi: 10.1136/amiajnl-2014-003144] [Medline: 25352570]

44.   Working for Quality. 2011. 2011 report to Congress: National strategy for quality improvements in healthcare URL: http://www.ahrq.gov/workingforquality/nqs/nqs2011annlrpt.htm [accessed 2015-06-25] [WebCite Cache ID 6ZYK33ism]

45.   Schnipper JL, Gandhi TK, Wald JS, Grant RW, Poon EG, Volk LA, et al. Effects of an online personal health record on medication accuracy and safety: a cluster-randomized trial. J Am Med Inform Assoc 2012;19(5):728-734 [FREE Full text] [doi: 10.1136/amiajnl-2011-000723] [Medline: 22556186]

46.   Wright A, Poon EG, Wald J, Feblowitz J, Pang JE, Schnipper JL, et al. Randomized controlled trial of health maintenance reminders provided directly to patients through an electronic PHR. J Gen Intern Med 2012 Jan;27(1):85-92 [FREE Full text] [doi: 10.1007/s11606-011-1859-6] [Medline: 21904945]

47.   Bhavnani V, Fisher B, Winfield M, Seed P. How patients use access to their electronic GP record--a quantitative study. Fam Pract 2011 Apr;28(2):188-194 [FREE Full text] [doi: 10.1093/fampra/cmq092] [Medline: 21084568]

48.   Delbanco T, Walker J, Bell SK, Darer JD, Elmore JG, Farag N, et al. Inviting patients to read their doctors' notes: a quasi-experimental study and a look ahead. Ann Intern Med 2012 Oct 2;157(7):461-470 [FREE Full text] [doi: 10.7326/0003-4819-157-7-201210020-00002] [Medline: 23027317]

49.   de Jong CC, Ros WJ, Schrijvers G. The effects on health behavior and health outcomes of Internet-based asynchronous communication between health providers and patients with a chronic condition: a systematic review. J Med Internet Res 2014;16(1):e19 [FREE Full text] [doi: 10.2196/jmir.3000] [Medline: 24434570]

50.   Lincoln YS, Guba EG. But is it rigorous? Trustworthiness and authenticity in naturalistic evaluation. New Directions for Program Evaluation 1986 ;1986(30):73-84. [doi: 10.1002/ev.1427]

Abbreviations

CADRE: Comprehensive Access and Delivery Research and Evaluation
CCD: Continuity of Care Document
EHR: electronic health records
HIE: health information exchanges
PHR: personal health records
VA: Veterans Affairs

Original Paper

Apps for IMproving FITness and Increasing Physical Activity Among Young People: The AIMFIT Pragmatic Randomized Controlled Trial

Artur Direito[1], MSc; Yannan Jiang[1], PhD; Robyn Whittaker[1], PhD; Ralph Maddison[1], PhD

Faculty of Medical and Health Sciences, National Institute for Health Innovation, University of Auckland, Auckland, New Zealand

Corresponding Author:
Artur Direito, MSc
Faculty of Medical and Health Sciences
National Institute for Health Innovation
University of Auckland
National Institute for Health Innovation, School of Population Health, University of Auckland
Private Bag 92019, Auckland Mail Centre
Auckland, 1142
New Zealand
Phone: 64 9 373 7599 ext 85285
Fax: 64 9 373 1710
Email: a.direito@auckland.ac.nz

Abstract

Background: Given the global prevalence of insufficient physical activity (PA), effective interventions that attenuate age-related decline in PA levels are needed. Mobile phone interventions that positively affect health (mHealth) show promise; however, their impact on PA levels and fitness in young people is unclear and little is known about what makes a good mHealth app.

Objective: The aim was to determine the effects of two commercially available smartphone apps (Zombies, Run and Get Running) on cardiorespiratory fitness and PA levels in insufficiently active healthy young people. A second aim was to identify the features of the app design that may contribute to improved fitness and PA levels.

Methods: Apps for IMproving FITness (AIMFIT) was a 3-arm, parallel, randomized controlled trial conducted in Auckland, New Zealand. Participants were recruited through advertisements in electronic mailing lists, local newspapers, flyers posted in community locations, and presentations at schools. Eligible young people aged 14-17 years were allocated at random to 1 of 3 conditions: (1) use of an immersive app (Zombies, Run), (2) use of a nonimmersive app (Get Running), or (3) usual behavior (control). Both smartphone apps consisted of a fully automated 8-week training program designed to improve fitness and ability to run 5 km; however, the immersive app featured a game-themed design and narrative. Intention-to-treat analysis was performed using data collected face-to-face at baseline and 8 weeks, and all regression models were adjusted for baseline outcome value and gender. The primary outcome was cardiorespiratory fitness, objectively assessed as time to complete the 1-mile run/walk test at 8 weeks. Secondary outcomes were PA levels (accelerometry and self-reported), enjoyment, psychological need satisfaction, self-efficacy, and acceptability and usability of the apps.

Results: A total of 51 participants were randomized to the immersive app intervention (n=17), nonimmersive app intervention (n=16), or the control group (n=18). The mean age of participants was 15.7 (SD 1.2) years; participants were mostly NZ Europeans (61%, 31/51) and 57% (29/51) were female. Overall retention rate was 96% (49/51). There was no significant intervention effect on the primary outcome using either of the apps. Compared to the control, time to complete the fitness test was –28.4 seconds shorter (95% CI –66.5 to 9.82, P=.20) for the immersive app group and –24.7 seconds (95% CI –63.5 to 14.2, P=.32) for the nonimmersive app group. No significant intervention effects were found for secondary outcomes.

Conclusions: Although apps have the ability to increase reach at a low cost, our pragmatic approach using readily available commercial apps as a stand-alone instrument did not have a significant effect on fitness. However, interest in future use of PA apps is promising and highlights a potentially important role of these tools in a multifaceted approach to increase fitness, promote PA, and consequently reduce the adverse health outcomes associated with insufficient activity.

XSL•FO
RenderX

Introduction

Worldwide, 80.3% (95% CI 80.1-80.5) of adolescents aged 13 to 15 years do not achieve current physical activity (PA) recommendations [1]. New Zealand mirrors these data with 67.7% in the age group of 15 to 19 years not complying with these guidelines [2]. A recent analysis of 50 studies including more than 25 million fitness test results of young people aged 9 to 17 years from 28 countries reported global declines of cardiorespiratory fitness (CRF) of approximately 5% every decade from 1970 to 2000 [3]. Insufficient activity is associated with increased risk of chronic diseases, such as cardiovascular diseases, type 2 diabetes mellitus, some types of cancer, and premature death [4,5]. Likewise, lower levels of CRF, an important marker of overall physiologic health [6,7] sensitive to patterns of PA, also predict future morbidity and mortality [8,9]. Increasing levels of PA and CRF in young people are important public health priorities for enhancing health, well-being, and preventing disease [10].

Physical activity declines with age throughout adolescence [2,11,12] and tracks into adulthood [13], making this an important time to intervene. Common approaches for intervening in this population have included school-based programs because schools integrate almost all young people. A Cochrane systematic review of school-based PA interventions targeting areas such as curriculum, teacher training, educational materials, and accessibility to exercise equipment (44 randomized controlled trials [RCTs], N=36,593, age 6-18 years) provided some evidence of effectiveness for increasing time spent in moderate-to-vigorous intensity PA (MVPA) from 5 to 45 minutes more per day and CRF (VO_{2peak}) from 1.6 to 3.7 mL/kg/min [14]. However, caution interpreting these findings is warranted because the included studies were at moderate risk of bias, with the majority having relied on self-reported activity [15]. Notably, a 2012 meta-analysis (30 RCTs, N=14,326, age ≤16 years), which differs from previous systematic reviews [16] in that it only included studies with accelerometry-measured outcomes (ie, objectively measured), showed small treatment effects for both total PA (standardized mean difference 0.12, 95% CI 0.04-0.20, *P*<.01) and MVPA (standardized mean difference 0.16, 95% CI 0.08-0.24, *P*<.001) [17]. These results translate into approximately 4 minutes more walking or running per day, a modest increase with limited clinical significance.

Many existing interventions, including school-based interventions, are limited because they are resource intensive. Young people are considered a hard-to-reach group and typically have limited adherence or exposure to PA interventions [16]. Therefore, novel approaches are crucial to engage those who are insufficiently active to encourage good health [18]. Given young peoples' ubiquitous use of mobile phones and increased digital literacy [19,20], interventions leveraging this technology may provide a promising intervention for this population.

Young people are early adopters of new consumer technology, such as smartphones and apps. Smartphone ownership among young people is on the rise, having increased from 23% in 2011 to 37% in 2012 in the United States, with similar distribution across ethnicity and family income [21]. Moreover, 58% of all US young people aged 12 to 17 years reported having downloaded apps compared to 34% of adults [22]. Likewise, digital divides have decreased between 2007 and 2013 in New Zealand [23], with 75% of high-income families reporting household access to a smartphone compared with 52% for those with lower income [24]. Alongside the increased ownership of smartphones, there has been a proliferation of apps in the major platform operators' app stores, which currently offer more than 100,000 mHealth apps [25]. Consumers looking for help via their smartphone face an immense number of mHealth apps [26] and the majority of these (31%) are fitness apps [25]. This high engagement with mobile technology offers an ideal opportunity to leverage the benefits of mobile interventions for health (mHealth), including lower participant burden and flexibility [27].

The effectiveness of mHealth-delivered interventions to promote PA that use texting or short message service (SMS) [28,29], personal digital assistants (PDAs) [30], and apps in addition to other components [31] has been examined. Unlike SMS text messaging, which has a substantial body of literature supporting its use, more complex interventions that capitalize on the computational power of smartphones have only recently been developed. Particularly, few studies have used an entirely mHealth device-based approach to deliver health behavior change interventions [32]. Most [33,34] include other components, such as a website, individual face-to-face or group sessions, phone calls, print materials, or pedometers, making it impossible to tease out the specific effects of each component. Thus, at present, there is a lack of scientific evidence assessing publicly available apps to promote PA. Because these apps are commercially available to the public on app stores and because fitness apps are currently the most popular (78% of users in 2014 compared to 39% in 2013 [35]), it is important to determine whether popular/commonly downloaded apps are effective [36].

Therefore, the primary aim of the Apps for IMproving FITness (AIMFIT) trial was to evaluate the effectiveness of two popular commercially available smartphone apps [37] for improving CRF in young people aged 14 to 17 years compared to usual

behavior alone (the control). Secondary aims were to determine the effect on PA levels, enjoyment, psychological need satisfaction, and self-efficacy. Perceptions of usability and acceptability of the apps were also assessed.

Methods

Trial Design

A 3-arm parallel RCT was conducted in Auckland, New Zealand. Details on the rationale, design, and methods have been previously described [38]. Ethics approval was obtained from the University of Auckland Human Participants Ethics Committee (10054/2013) and the study was conducted according to the principles of the Declaration of Helsinki. There were no deviations to methods after trial commencement. All participants (and guardians for participants younger than 16 years) provided written informed consent. The trial was registered (Australian New Zealand Clinical Trials Registry: ACTRN12613001030763) in accordance with the requirements of the International Committee of Medical Journal Editors and reported in accordance with the CONSORT statement [39] and the CONSORT-EHEALTH extension [40] (see Multimedia Appendix 1). A brief description of the procedures is provided.

Participants

Participants were recruited through advertisements in electronic mailing lists, local newspapers, schools, and flyers posted in community locations. Consenting schools and churches allowed the researcher to present a brief outline of the study. Those interested provided contact details and their eligibility was assessed via telephone. If eligible, participant information and informed consent documentation were either mailed or emailed and participants were scheduled to attend a face-to-face baseline assessment at the university.

Eligible participants were aged 14 to 17 years, lived in Auckland, owned an iPod touch or smartphone running at least Android 2.2 or iOS 6.0, and were able to perform physical activities but were not achieving [41] the PA recommendations (ie, at least 60 minutes of MVPA each day). Exclusion criteria were a medical condition limiting ability to exercise safely, previous use of the apps of interest, and inability to comply with the study protocol. Only 1 child per household was eligible to take part.

Randomization

Participants were enrolled by author AD and were randomly assigned at a 1:1:1 ratio to 1 of 3 conditions. Stratified block randomization in variable blocks was used to maintain balance across gender, an important prognostic factor [2,42]. A biostatistician (YJ) prepared the randomization scheme in advance by using a computer-generated randomization table. Based on the randomization scheme, a research assistant prepared opaque-sealed envelopes containing group referral so that the researcher could not identify group assignment. The envelope was opened by each participant after completion of baseline assessment. Given the nature of the intervention, it was not possible to blind participants. However, allocation concealment was maintained up to the point of randomization.

Interventions

Commercially available apps targeting fitness were identified during previous work evaluating the most popular (ie, top-20 free and top-20 paid) downloaded apps in the Health and Fitness Category of the iTunes New Zealand store [37]. The control group was compared against 2 other groups, each using an app to promote fitness available on both of the major platforms' app stores (iTunes and Google Play Store). The 2 intervention groups allowed comparison of an identical delivery approach (ie, stand-alone app), but with distinct design features: (1) use of an immersive app or (2) use of a nonimmersive app. Both apps consisted of a fully automated 8-week training program designed to improve fitness and ability to run 5 km; however, the immersive app featured a game-themed design whereby the training program was embedded with a story where the user is trained to collect supplies and protect a town from zombies. Effective self-regulatory behavior change techniques [43,44] underlie the apps, which allow the users to self-monitor and receive feedback on their training. The apps provided information on running and technique, audio instructions on how to perform the training components, and tracked and displayed progress throughout the program. Other features of the apps include the ability to work out with music on the device's library and links to associated websites to interact with other users.

Participants randomized to the immersive app group received the Zombies, Run! 5K Training app developed by Six to Start with Naomi Alderman for iOS and the Android platform. It was released worldwide for iOS on October 2012. Even though data on number of downloads is not publicly available, the Google Play Store reports 100,000 to 500,000 installs of this app, and the Zombies, Run! community has more than 800,000 players worldwide [45]. Those randomized to the nonimmersive app group received the Get Running-Couch to 5k app, developed by Splendid Things and Benjohn Barnes for iOS and the Android platform. It was released worldwide for Android on July 2009 and the Google play Store reports 10,000 to 50,000 installs of this app [46]. While the trial was conducted, the apps cost NZ $2.45 to $4.19, depending on the app store. Given the commercialized nature of the apps, updates occurred during the trial; however, no major changes affected their content or the intervention. Videos detailing the features of the apps can be accessed via links on the preceding references.

Following randomization, the respective app was paid for and installed by AD on each participant's mobile device and a short instruction on the features and settings of the app was given. Participants were encouraged to use their app 3 times per week and work their way through each of the workouts, but because this was a pragmatic study [47], access and usage was allowed to vary (ie, participants were able to use at their own pace, ad libitum, as it would occur on an "everyday life" setting). There were no cointerventions, no supplementary modes of delivery, nor usage of prompts (ie, emails, phone calls, SMS text message) to use the app.

Control Group

The control group was asked to continue with their usual physical activities for the duration of the study and was not

offered any information about increasing PA. Both apps were provided (free of charge) to participants after trial completion.

Procedure

Assessments were conducted at baseline and 8 weeks at the university by AD. Participants were assessed individually. At both time points, participants completed a field test of CRF (1 mile run/walk test), had their height and weight measured, self-reported their physical activity and related psychological variables, were given an Actigraph accelerometer to wear for the following 7 days (to provide an objective assessment of their free-living PA), and completed a booklet detailing their accelerometer use. AD collected the accelerometers and booklets from the participants' homes (during the randomization visit that took place after the baseline assessment and at the last visit after the follow-up assessment). Follow-up assessments were not blinded. Participants received a NZ $10 gift card to a local shopping center for each visit to complete study measures (ie, maximum NZ $30 for 3 visits). The vouchers were not conditional on usage of the app; they were offered to compensate for participants' time and encourage completion of study measures.

Outcomes

The primary outcome was CRF, assessed with the 1-mile run/walk test. Following the procedures outlined in the Fitnessgram test administration manual, participants were instructed to run and/or walk at their own pace until completing the distance in the shortest possible time [48]. The output, in seconds, was used to estimate VO_{2peak} using a validated prediction equation [49].

Secondary outcomes included anthropometrics, self-reported PA and associated psychological variables, objectively measured PA, and self-reported acceptability and usability of the apps assessed via an exit survey conducted with intervention participants. A series of closed and open-ended questions were asked to determine features perceived as more and less acceptable as well as which features participants found more useful to support their fitness. Body weight (in kg, without shoes) was measured with a Salter scale to 1 decimal place. Height was measured to the nearest 0.1 cm with a Seca stadiometer. Two measurements were taken for each and the means were used for analysis. Body mass index (BMI) was calculated by using the standard equation (weight in kilograms/height in meters squared). BMI-for-age was calculated using the World Health Organization (WHO) growth standards macro [50].

Using instruments validated in this population, participants self-reported (1) physical activity using the Physical Activity Questionnaire for Adolescents (PAQ-A) [51,52], (2) perceived enjoyment using the Physical Activity Enjoyment Scale (PACES) [53,54], (3) perceived competence, autonomy, and relatedness using the Psychological Need Satisfaction in Exercise Scale (PNSES) [55], (4) self-efficacy using the Physical Activity Self-Efficacy Scale (PASES) [56,57], and (5) frequency of app utilization, acceptability, and usability of the app via an exit survey conducted with the participants in the intervention groups.

Participants were instructed to wear the accelerometer (Actigraph GT1M) on their right hip during waking hours for 7 days after each assessment, removing it when engaging in activities involving water and/or contact sports. A 10-second epoch was used and data were aggregated into minute intervals for subsequent processing. To determine valid wear time, periods of more than 60 minutes of consecutive zeroes and days with less than 600 minutes of valid records were removed before data analysis [58]. Participants were included in the analyses if they had 3 or more valid monitoring days [59]. Mean daily time (min) spent in sedentary and light-to-vigorous activities was calculated for each participant using the cut points proposed by Evenson et al to categorize intensities [60,61].

Adverse events were collected at each study visit or voluntarily reported by contacting the researcher. An adverse event was considered serious if it required hospitalization.

Sample Size Calculation

A total of 51 participants (17 per group) was estimated to provide 80% power and α=.05 overall to detect a difference of 17 seconds in CRF, assuming a 15-second SD in time to complete the 1-mile run/walk test between each of the conditions compared to the control [62].

Statistical Analyses

Treatment evaluations were performed on the principle of intent-to-treat, including all randomized participants as allocated. Statistical analyses were performed with SAS version 9.4 software (SAS Institute, Cary, NC, USA). All statistical tests were 2-sided at a 5% significance level, with adjustment for multiple comparisons on the primary outcome. Analysis of covariance (ANCOVA) regression model was used to evaluate the main treatment effects on the primary outcome, adjusting for baseline measure and gender. Model-adjusted means, 95% confidence intervals, and *P* values were estimated for each group. Mean differences between groups were tested. A similar approach was used for secondary outcomes. Missing data on the primary outcome were imputed with the baseline value. Per-protocol analyses were conducted on those participants with complete baseline and follow-up data and self-reported adherence to the training program (ie, app used 3 times/week) following the same procedures as in the intention-to-treat analyses.

Results

Overview

Recruitment began October 2013 and finished in June 2014. The final follow-up visit was in September 2014. Figure 1 presents the participant flow diagram. Of 143 individuals screened, 51 eligible participants were randomized to the Zombies, Run immersive app intervention (n=17), the Get Running nonimmersive app intervention (n=16), or the control group (n=18). Baseline demographic characteristics of participants are presented in Table 1.

Participants had a mean age of 15.7 years (SD 1.2, range 14-17 years) and a BMI of 22.9 (SD 4.3) kg/m². The majority were NZ European (61%, 31/51), whereas 22% (11/51) were Pacific

Islanders, and 57% (29/51) were female. Follow-up assessments at 8 weeks were completed for 17 (100%, 17/17) immersive app group participants, 15 (94%, 15/16) nonimmersive app group participants, and 17 (94%, 17/18) control group participants, which represents an overall retention rate of 96% (49/51) from baseline.

Table 1. Baseline demographic and clinical characteristics.

Characteristic	Zombies, Run n=17	Get Running n=16	Control n=18	Total N=51
Age (years), mean (SD)	15.78 (1.11)	15.69 (1.04)	15.55 (1.32)	15.67 (1.15)
Gender, n (%)				
Male	8 (47)	6 (38)	8 (44)	22 (43)
Female	9 (53)	10 (63)	10 (56)	29 (57)
Ethnicity, n (%)				
Maori	3 (18)	0 (0)	0 (0)	3 (6)
NZ European	9 (53)	9 (56)	13 (72)	31 (61)
Pacific	4 (24)	3 (19)	4 (22)	11 (22)
Asian	0 (0)	3 (19)	1 (6)	4 (8)
Other	1 (6)	1 (6)	0 (0)	2 (4)
BMI (kg/m^2), mean (SD)	23.17 (3.60)	21.85 (3.14)	23.43 (5.56)	22.85 (4.25)
BMI-for-age[a] (z-score), mean (SD)	0.77 (0.86)	0.36 (0.93)	0.64 (1.46)	0.60 (1.12)
Device, n (%)				
iPhone	8 (47)	6 (38)	11 (61)	25 (49)
Android	5 (29)	7 (44)	5 (28)	17 (33)
iPod Touch	4 (24)	3 (19)	2 (11)	9 (18)
Estimated VO_{2peak}[b] (mL/kg/min), mean (SD)	43.51 (6.11)	43.58 (5.47)	44.20 (6.95)	43.78 (6.12)

[a] WHO growth reference.

[b] Prediction equation from 1-mile run/walk test.

Figure 1. Flow diagram of the Apps for IMproving FITness (AIMFIT) trial. Those who were unable to complete the postintervention fitness assessment due to injury or sickness still completed self-reported outcomes and were included in all analyses.
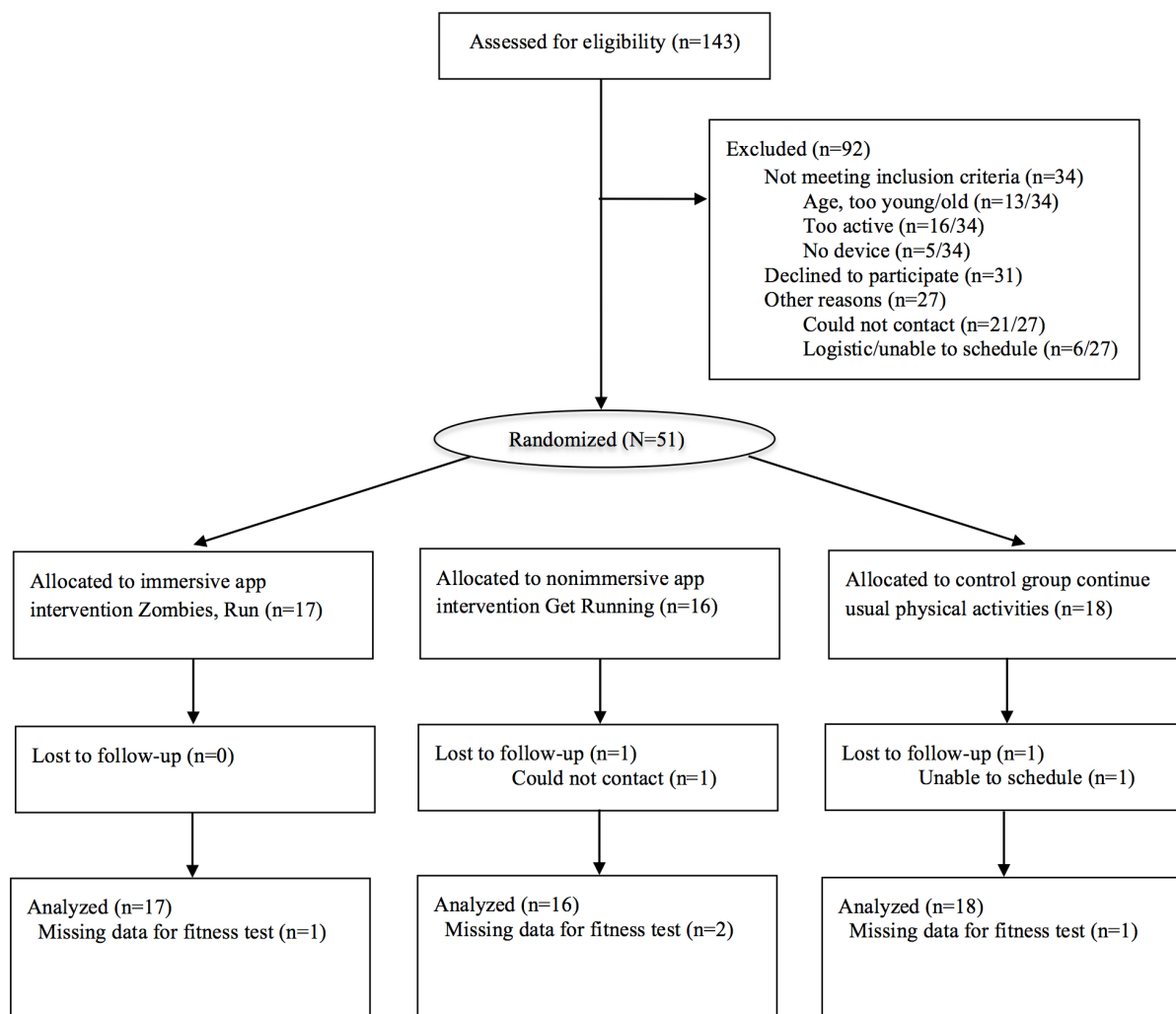
Primary Outcome

Table 2 lists the effects of the app interventions on the time to complete the 1-mile walk/run fitness test and all secondary outcomes at 8 weeks. On average, time to complete the fitness test decreased in both app groups, but there were no statistically significant differences observed between the intervention groups and the control (immersive app group: adjusted mean difference –28.4 sec, 95% CI –66.5 to 9.8, P=.20; nonimmersive app group: adjusted mean difference –24.7 sec, 95% CI –63.5 to 14.2 sec, P=.32). To facilitate interpretation of test scores, estimated $VO_{2\,peak}$ [49] at 8 weeks was 44.09, 44.59, and 43.44 mL/kg/min for the immersive app, nonimmersive app, and control groups, respectively.

Table 2. Treatment effects at 8 weeks.

Outcome	Zombies, Run (1), mean (SD)		Get Running (2), mean (SD)		Control (3), mean (SD)		Adjusted difference[a] (95% CI) at 8 weeks			
	Base-line	8 week	Base-line	8 week	Base-line	8 week	1 vs 3	P	2 vs 3	P
Time to complete 1-mile walk/run (sec)	574.06 (145.68)	560.06 (139.27)	586.56 (129.74)	576.75 (147.91)	585.89 (600.17)	600.17 (191.38)	–28.36 (–66.54, 9.82)	.20	–24.67 (–63.51, 14.18)	.32
PAQ-A	2.20 (0.66)	2.27 (0.53)	2.09 (0.73)	2.31 (0.74)	2.30 (0.67)	2.21 (0.62)	0.14 (–0.26, 0.54)	.78	0.23 (–0.18, 0.64)	.42
PACES	4.08 (0.47)	4.00 (0.46)	3.99 (0.46)	3.85 (0.46)	3.96 (0.58)	4.00 (0.57)	–0.10 (–0.33, 0.13)	.62	–0.17 (–0.40, 0.06)	.19
PNSES	4.52 (0.69)	4.49 (0.78)	4.48 (0.89)	4.56 (0.56)	4.67 (0.85)	4.68 (0.76)	–0.08 (–0.46, 0.31)	.95	0.01 (–0.38, 0.40)	>.99
Competence	4.27 (0.83)	4.24 (0.94)	4.25 (1.09)	4.32 (0.94)	4.54 (1.17)	4.52 (1.22)	–0.08 (–0.67, 0.51)	.98	0.03 (–0.57, 0.63)	.99
Autonomy	4.87 (0.98)	4.92 (0.94)	4.94 (0.78)	5.21 (0.56)	4.94 (0.99)	4.84 (0.91)	0.12 (–0.44, 0.69)	.93	0.36 (–0.22, 0.94)	.34
Relatedness	4.40 (0.79)	4.31 (0.87)	4.26 (1.15)	4.14 (0.91)	4.53 (1.16)	4.67 (1.27)	–0.26 (–0.88, 0.35)	.65	–0.34 (–0.97, 0.29)	.45
PASES	2.50 (0.32)	2.44 (0.40)	2.38 (0.47)	2.43 (0.31)	2.38 (0.41)	2.39 (0.34)	–0.02 (–0.24, 0.19)	.99	0.04 (–0.18, 0.26)	.96
Average daily valid use (min)	801.16 (87.85)	784.43 (93.69)	819.39 (85.45)	831.08 (78.74)	815.59 (57.14)	814.12 (73.61)	–21.72 (–84.00, 40.56)	.77	13.71 (–49.56, 76.99)	.93
Average daily activity counts (counts/min)	354.27 (123.75)	341.21 (146.22)	269.97 (82.72)	270.20 (84.77)	364.83 (153.81)	327.53 (140.02)	17.74 (–63.07, 98.55)	.93	0.90 (–85.41, 87.22)	>.99
Average daily time spent in sedentary activities (min)	526.63 (106.37)	535.22 (113.90)	553.64 (96.80)	570.63 (95.58)	529.43 (94.64)	548.18 (94.11)	–10.94 (–69.83, 48.00)	.96	3.95 (–56.26, 64.16)	.99
Average daily time spent in light PA (min)	237.41 (73.32)	216.17 (67.46)	244.46 (61.97)	237.11 (58.45)	250.39 (64.05)	235.40 (63.89)	–10.54 (–53.96, 32.88)	.91	4.12 (–39.94, 48.17)	.99
Average daily time spent in moderate PA (min)	30.28 (14.28)	25.07 (12.83)	17.03 (7.81)	18.03 (10.20)	25.52 (13.75)	22.33 (11.69)	1.42 (–7.96, 10.81)	.98	–1.71 (–11.51, 8.10)	.96
Average daily time spent in vigorous PA (min)	6.84 (6.03)	7.97 (9.16)	4.26 (4.21)	5.30 (4.34)	10.26 (11.72)	8.22 (9.22)	1.26 (–3.82, 6.33)	.90	0.52 (–4.79, 5.83)	.99
Average daily time spent in MVPA (min)	37.12 (16.84)	33.04 (20.61)	21.29 (11.25)	23.34 (14.04)	35.78 (22.54)	30.54 (17.99)	1.74 (–11.45, 14.93)	.98	–1.82 (–16.00, 12.36)	.99

[a] Adjusted for baseline, gender, and multiple comparisons.

Secondary Outcomes

No intervention effects were found for self-reported secondary outcomes of physical activity (PAQ-A; immersive app group: adjusted mean difference 0.14, 95% CI –0.26 to 0.54, P=.78; nonimmersive app group: adjusted mean difference 0.23, 95% CI –0.18 to 0.64, P=.42) or its predictors of perceived enjoyment (PACES), perceived competence, autonomy, and relatedness (PNSES), or self-efficacy (PASES) (see Table 2 for all outcomes).

For accelerometry, 48 of 51 (94%) participants provided valid data for analysis at baseline, whereas compliance with wearing the device slightly decreased at postintervention (46/51, 90%). Group assignment did not have a significant effect on overall activity (ie, mean counts per minute) or mean daily time spent in MVPA. Compared to the control group, mean baseline daily time spent in MVPA-, gender-, and multiple comparisons-adjusted time in MVPA difference was 1.74 min (95% CI –11.45 to 14.93, P=.98) and –1.82 min (95% CI –16.00 to 12.36, P=.99) for the immersive and nonimmersive app groups, respectively.

A total of 6 adverse events (1 serious) were reported in 6 participants, 4 of which were in the control group (ankle injury-2 events, lower back pain, and hospitalization because of tonsils removal) and 1 in each of the intervention groups (ankle injury-2 events). None of the adverse events were deemed related to the study intervention.

XSL•FO
RenderX

Approximately two-thirds of participants in the intervention groups reported using the app either 2 (10/32, 31%) or 3 times per week (10/32, 31%), whereas 8 of 32 (25%) only used it 1 time per week (see Multimedia Appendix 2). No differences were evident on timing of use (ie, weekday, weekend, morning, afternoon, evening). Apps were mostly used outdoors (eg, street, park vs gym, home treadmill) and while alone (n=13 alone, n=7 with friend, and n=9 with family).

For the app Zombies, Run!, the features mostly used by participants were the "workout mission tasks" (n=14) and "story and run log" of completed workouts (n=10), whereas social networking features ("share my runs": n=0; "ZombieLink account": n=3) were seldom or never used. Results were similar when participants reported the features they liked ("workout mission tasks": n=14; "story and run log": n=8) and disliked ("share my runs": n=5; "ZombieLink account": n=5).

For the app Get Running, the feature mostly used by participants was the description of the "week-runs" (n=13), whereas only 1 participant reported using the social networking feature "status updates." The description of the "week-runs" was also the feature participants predominantly liked (n=11), whereas the main feature disliked was the "status updates" (n=3).

Regardless of the app used, similar themes emerged when participants reported their willingness and motives to continue using their app after study participation. Those willing to continue stated personal benefits (eg, "It will help me to build my fitness", "Because I can improve how far I run") and app-related motives (eg, "A fun way to get fit", "Because it is an enjoyable alternative to exercise"). For those unwilling to continue, "not enough time" was the most common barrier, followed by lack of interest (eg, "I didn't find the app engaging enough"). The nonimmersive app received less positive feedback around motivational aspects (eg, "Using the app became too tedious").

Overall, participants perceived the layout of the apps and menus as well structured and "straightforward" to use. Being able to receive clear instructions (eg, "Tells me what to do and when"), listen to their own music during the training sessions, task difficulty increasing gradually, and encouragement provided were features highlighted as useful to support participants' fitness. Some also considered it helpful if the app allowed choosing between different goals and activities (eg, "I prefer to run to my own goals"). For the immersive app, the storyline ("The back story made it interesting") and the ability to track progress (ie, app used the device's Global Positioning System [GPS] and/or accelerometer to log distance) (eg, "It records the distance you ran and your time so you are able to view it for next time and compare") were also reported as important features.

The majority of participants (21/32, 66%) had no prior experience of using their smartphone for PA purposes. Examples of prior experience included listening to music while engaging in PA or previous use of free apps (eg, MyFitnessPal). Overall, 81% (26/32) were interested in trying different PA-promoting apps in the future.

In prespecified per-protocol analyses (ie, the app was used 3 times/week), there were statistically significant differences observed on the primary outcome between the nonimmersive app group and the control (adjusted mean difference –79.39 sec, 95% CI –133.01 to –25.77, P=.003). Time to complete the fitness test also decreased for the immersive app group compared to the control, with the difference not meeting statistical significance (adjusted mean difference –55.29 sec, 95% CI –111.46 to 0.88, P=.06). No intervention effects were found for self-reported secondary outcomes of PA, its predictors, or accelerometry.

Discussion

Principal Findings

This is the first randomized trial comparing the effects of a stand-alone immersive mobile app and a nonimmersive app on CRF, PA levels, and its predictors in young people. Key findings were that fitness improved in both app groups, but these did not significantly differ from the control. Despite the availability of readily available commercial apps to improve health behavior, these findings suggest that, compared to usual care, no major improvements were found for these 2 top downloaded apps.

The small increases in fitness in the present trial (0.6 to 1.0 mL/kg/min) were lower than those observed in a Cochrane review of school-based PA interventions, which found increases of 1.6 to 3.7 mL/kg/min in VO_{2peak} [14]. Similarly, Baquet et al [63], who reviewed the effects of training programs in young people's aerobic fitness, showed improvements on average of 5% to 6% in VO_{2peak}, whereas fitness only improved 1.3% to 2.3% in this study. The training programs of the apps were consistent with the frequency and duration of programs shown to improve VO_{2peak} [63] (ie, incorporating 3 sessions per week, each lasting 30 minutes to 1 hour). However, use of the apps during the AIMFIT study was ad libitum and not closely monitored, which is how app use would likely occur in real life. Smaller fitness improvements in the current study can most likely be attributed to lower adherence to app use (only 31.3% reported using 3 times/week) or due to participants not having exercised at a sufficient intensity to produce larger fitness effects [63], as observed by no changes in accelerometry-measured MVPA.

Although apps have the potential to increase the reach of health behavior change interventions, our results mirror recent research highlighting that only some participants will consistently use an offered app (approximately 20%) [64] or sustain use over time (approximately 3%) [65]. App design, features, and usability are important factors that impact users' engagement with an app. Although in AIMFIT both app groups produced comparable fitness effects, the design and features of the immersive app received more positive feedback (and no dropout) and, therefore, these aspects should be considered for future app development. Although typical app use is intermittent [65] and unlikely to induce the more intensive type of immersion observed in other videogame play, incorporating gaming elements that entangle the virtual world (ie, leisure) with actual

experience (ie, exertion) could potentially increase user engagement and sustained use.

An important consideration of app content is whether or not they incorporate behavior change techniques (BCTs). Further, modeling, providing consequences for behavior, providing information on others' approval, prompting intention formation, self-monitoring, and a behavioral contract were identified as effective BCTs for increasing PA in young people in a recent meta-regression [66]. A notable exception is providing instruction, which was associated with decreased effectiveness of PA interventions in young people, but has been shown to be highly present in top-ranked PA apps [37], including the ones hereby investigated. Providing instructions and not incorporating other effective BCTs or incorrectly applying BCTs (ie, not considering the parameters for them to be effective) may have undermined app efficacy.

Consistent with the primary outcome findings, we found no changes in any of the measured psychological variables. Fulfillment of the 3 basic psychological needs (ie, autonomy, competence, and relatedness)—key elements in the development of intrinsic motivation required to drive behavioral change [67]—did not change. This could be explained by the nature of the training programs, which were based on provision of instructions and may have been perceived as too prescriptive and controlling. Perhaps the alternatives provided by the apps (ie, walking, jogging, sprinting, running) lacked choice and encouragement of the participants' initiative and, therefore, did not support their autonomy. In addition, skill acquisition and feelings of competence did not change as a function of the intervention. This suggests the BCTs in the apps were insufficient to manipulate these variables or that app use was not intensive enough to evoke change. Further, the apps incorporate features that could capitalize on relatedness and social support, such as online communities, but some of these features load outside of the apps (eg, ZombieLink), which may have contributed to poor use of such features. Interestingly, enjoyment, which is associated with adherence to PA programs, did not change.

Strengths and Limitations

A major strength of AIMFIT was the use of a RCT design to determine the effectiveness of 2 off-the-shelf commercially available interventions. We chose a pragmatic approach in which participants used their own device and apps were used ad libitum. Contact with participants was minimal, which reflects app use in a real-world context and therefore increases the generalizability of the findings. Moreover, the primary outcome was assessed objectively with a valid and reliable measurement, as well as PA via accelerometry, which adds to the study's internal validity. Unlike the commonly observed high attrition in eHealth and mHealth interventions, follow-up assessments were completed for 96% of participants, which represents a high retention rate. Further, using paid apps instead of free ones has likely reduced the chances of contamination and/or cointerventions. Although the accessibility of these apps on the app stores could have led participants to download them, our study sample was a young population and the apps under investigation required purchase. Although inexpensive, purchasing an app on an app store requires a credit card, which young people do not typically own.

The major limitation of this study was its low statistical power and small sample size. We based our sample size calculation on a smaller standard deviation of the primary outcome than the actual standard deviation observed, which meant the power of the trial was smaller than 80%. We used readily available apps and consequently were limited to the decisions made by the app developers on content, duration of the program, and design features. This also meant that we were unable to access data on app utilization (eg, menus accessed in the app). Further, the relatively short duration of the programs precluded investigation of long-term effects or sustainability. This study also highlights that the peer-reviewed literature will always lag behind consumer technology life cycles because during the lifetime of this study innovative apps were developed at a rate that far outpaced our capacity to test them.

Implications

Among app users, fitness apps are the most popular (78% users in 2014 compared to 39% in 2013) [35]. However, off-the-shelf common commercial mobile apps used as a stand-alone approach to change fitness and promote PA in insufficiently active individuals do not appear effective. Future use as part of a multifaceted approach to increase fitness and to promote PA may be useful. The current one-size-fits-all approach of most common commercial mobile apps is limiting for both users and researchers. More tailored approaches, which are dynamic and responsive to changes in PA behavior over time, are required to improve the ease of use of apps, user engagement, and the apps' sustained use.

Conclusions

Readily available commercial apps as a stand-alone intervention to improve fitness and increase PA in young people did not increase CRF compared to usual care. Given that smartphone technology appears to resonate with young people and that this type of self-guided intervention has the potential to increase reach at a low cost, this may be best suited as part of a multicomponent intervention, providing additional support and encouragement to the participants (eg, maintenance phases).

XSL•FO

RenderX

Authors' Contributions

AD and RM conceived the study, participated in its design and coordination, and helped draft the manuscript. YJ and RW participated in the design of the study and helped draft the manuscript. All authors read and approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1

CONSORT-EHEALTH checklist V1.6.2 [40].

[PDF File (Adobe PDF File), 146KB - jmir_v17i8e210_app1.pdf]

Multimedia Appendix 2

Supplementary word file containing graphics with exit survey questions for intervention participants.

[PDF File (Adobe PDF File), 417KB - jmir_v17i8e210_app2.pdf]

References

1. Hallal PC, Andersen LB, Bull FC, Guthold R, Haskell W, Ekelund U, Lancet Physical Activity Series Working Group. Global physical activity levels: surveillance progress, pitfalls, and prospects. Lancet 2012 Jul 21;380(9838):247-257. [doi: 10.1016/S0140-6736(12)60646-1] [Medline: 22818937]

2. Clinical Trials Research Unit, Synovate. A National Survey of Children and Young People's Physical Activity and Dietary Behaviours in New Zealand: 2008/09: Key Findings. Auckland, New Zealand: Clinical Trials Research Unit & Synovate; 2010.

3. Tomkinson GR, Annandale M, Ferrar K. Global changes in cardiovascular endurance of children and youth since 1964: systematic analysis of 25 million fitness test results from 28 countries. Circulation 2013;128(22 Supplement):A13498.

4. Lee I, Shiroma EJ, Lobelo F, Puska P, Blair SN, Katzmarzyk PT, Lancet Physical Activity Series Working Group. Effect of physical inactivity on major non-communicable diseases worldwide: an analysis of burden of disease and life expectancy. Lancet 2012 Jul 21;380(9838):219-229 [FREE Full text] [doi: 10.1016/S0140-6736(12)61031-9] [Medline: 22818936]

5. Ekelund U, Luan J, Sherar LB, Esliger DW, Griew P, Cooper A, International Children's Accelerometry Database (ICAD) Collaborators. Moderate to vigorous physical activity and sedentary time and cardiometabolic risk factors in children and adolescents. JAMA 2012 Feb 15;307(7):704-712 [FREE Full text] [doi: 10.1001/jama.2012.156] [Medline: 22337681]

6. Kaminsky LA, Arena R, Beckie TM, Brubaker PH, Church TS, Forman DE, American Heart Association Advocacy Coordinating Committee, Council on Clinical Cardiology,Council on Nutrition, Physical ActivityMetabolism. The importance of cardiorespiratory fitness in the United States: the need for a national registry: a policy statement from the American Heart Association. Circulation 2013 Feb 5;127(5):652-662 [FREE Full text] [doi: 10.1161/CIR.0b013e31827ee100] [Medline: 23295916]

7. Ortega FB, Ruiz JR, Castillo MJ, Sjöström M. Physical fitness in childhood and adolescence: a powerful marker of health. Int J Obes (Lond) 2008 Jan;32(1):1-11. [doi: 10.1038/sj.ijo.0803774] [Medline: 18043605]

8. Shuval K, Finley CE, Barlow CE, Gabriel KP, Leonard D, Kohl HW. Sedentary behavior, cardiorespiratory fitness, physical activity, and cardiometabolic risk in men: the cooper center longitudinal study. Mayo Clin Proc 2014 Aug;89(8):1052-1062. [doi: 10.1016/j.mayocp.2014.04.026] [Medline: 25034308]

9. Kodama S, Saito K, Tanaka S, Maki M, Yachi Y, Asumi M, et al. Cardiorespiratory fitness as a quantitative predictor of all-cause mortality and cardiovascular events in healthy men and women: a meta-analysis. JAMA 2009 May 20;301(19):2024-2035. [doi: 10.1001/jama.2009.681] [Medline: 19454641]

10. Marques A, Santos R, Ekelund U, Sardinha LB. Association between physical activity, sedentary time, and healthy fitness in youth. Med Sci Sports Exerc 2015 Mar;47(3):575-580. [doi: 10.1249/MSS.0000000000000426] [Medline: 24977696]

11. Dumith SC, Gigante DP, Domingues MR, Kohl HW. Physical activity change during adolescence: a systematic review and a pooled analysis. Int J Epidemiol 2011 Jun;40(3):685-698 [FREE Full text] [doi: 10.1093/ije/dyq272] [Medline: 21245072]

12. Corder K, Sharp SJ, Atkin AJ, Griffin SJ, Jones AP, Ekelund U, et al. Change in objectively measured physical activity during the transition to adolescence. Br J Sports Med 2014 Apr 9;49(11):730-736 [FREE Full text] [doi: 10.1136/bjsports-2013-093190] [Medline: 24273308]

13. Telama R. Tracking of physical activity from childhood to adulthood: a review. Obes Facts 2009;2(3):187-195. [doi: 10.1159/000222244] [Medline: 20054224]

14. Dobbins M, Husson H, DeCorby K, LaRocca RL. School-based physical activity programs for promoting physical activity and fitness in children and adolescents aged 6 to 18. Cochrane Database Syst Rev 2013;2:CD007651. [doi: 10.1002/14651858.CD007651.pub2] [Medline: 23450577]

15. Taber DR, Stevens J, Murray DM, Elder JP, Webber LS, Jobe JB, et al. The effect of a physical activity intervention on bias in self-reported activity. Ann Epidemiol 2009 May;19(5):316-322 [FREE Full text] [doi: 10.1016/j.annepidem.2009.01.001] [Medline: 19230711]

16. van Sluijs EM, McMinn AM, Griffin SJ. Effectiveness of interventions to promote physical activity in children and adolescents: systematic review of controlled trials. BMJ 2007 Oct 6;335(7622):703 [FREE Full text] [doi: 10.1136/bmj.39320.843947.BE] [Medline: 17884863]

17. Metcalf B, Henley W, Wilkin T. Effectiveness of intervention on physical activity of children: systematic review and meta-analysis of controlled trials with objectively measured outcomes (EarlyBird 54). BMJ 2012;345:e5888 [FREE Full text] [Medline: 23044984]

18. Hallal P, Bauman A, Heath G, Kohl H, Lee I, Pratt M. Physical activity: more of the same is not enough. The Lancet 2012;380(9838):190-191.

19. Kontos E, Blake KD, Chou WS, Prestin A. Predictors of eHealth usage: insights on the digital divide from the Health Information National Trends Survey 2012. J Med Internet Res 2014;16(7):e172 [FREE Full text] [doi: 10.2196/jmir.3117] [Medline: 25048379]

20. Eshet-Alkalai Y, Chajut E. Changes over time in digital literacy. Cyberpsychol Behav 2009 Dec;12(6):713-715. [doi: 10.1089/cpb.2008.0264] [Medline: 19514816]

21. Madden M, Lenhart A, Duggan M, Cortesi S, Gasser U. Teens and Technology 2013. Washington, DC: Pew Internet & American Life Project; 2013. URL: http://www.pewinternet.org/2013/03/13/teens-and-technology-2013/ [accessed 2015-04-21] [WebCite Cache ID 6XxouIZKW]

22. Madden M, Lenhart A, Cortesi S, Gasser U. Teens and Mobile Apps Privacy. Washington, DC: Pew Internet & American Life Project; 2013. URL: http://www.pewinternet.org/2013/08/22/teens-and-mobile-apps-privacy/ [accessed 2015-04-21] [WebCite Cache ID 6XxpErD48]

23. Crothers C, Gibson A, Smith P, Bell A, Miller M. Internet Trends in New Zealand 2007–2013. Auckland, New Zealand: Institute of Culture, Discourse & Communication, Auckland University of Technology; 2014.

24. Crothers C, Gibson A, Smith P, Bell A, Miller M. The Internet in New Zealand 2013. Auckland, New Zealand: A Institute of Culture, Discourse & Communication, AUT University; 2013.

25. Reseach2Guidance. Mobile Health Economics. 2014. Mhealth app developer economics 2014-the state of the art of mhealth app publishing URL: http://mhealtheconomics.com/reports/ [accessed 2015-04-21] [WebCite Cache ID 6XxnIzwSd]

26. Kuehn BM. Is there an app to solve app overload? JAMA 2015 Apr 14;313(14):1405-1407. [doi: 10.1001/jama.2015.2381] [Medline: 25871657]

27. Tate EB, Spruijt-Metz D, O'Reilly G, Jordan-Marsh M, Gotsis M, Pentz MA, et al. mHealth approaches to child obesity prevention: successes, unique challenges, and next directions. Transl Behav Med 2013 Dec;3(4):406-415 [FREE Full text] [doi: 10.1007/s13142-013-0222-3] [Medline: 24294329]

28. Hurling R, Catt M, Boni M, Fairley B, Hurst T, Murray P, et al. Using internet and mobile phone technology to deliver an automated physical activity program: randomized controlled trial. J Med Internet Res 2007;9(2):e7 [FREE Full text] [doi: 10.2196/jmir.9.2.e7] [Medline: 17478409]

29. Fjeldsoe BS, Miller YD, Marshall AL. MobileMums: a randomized controlled trial of an SMS-based physical activity intervention. Ann Behav Med 2010 May;39(2):101-111. [doi: 10.1007/s12160-010-9170-z] [Medline: 20174902]

30. Conroy MB, Yang K, Elci OU, Gabriel KP, Styn MA, Wang J, et al. Physical activity self-monitoring and weight loss: 6-month results of the SMART trial. Med Sci Sports Exerc 2011 Aug;43(8):1568-1574 [FREE Full text] [doi: 10.1249/MSS.0b013e31820b9395] [Medline: 21200337]

31. Kirwan M, Duncan MJ, Vandelanotte C, Mummery WK. Using smartphone technology to monitor physical activity in the 10,000 Steps program: a matched case-control trial. J Med Internet Res 2012;14(2):e55 [FREE Full text] [doi: 10.2196/jmir.1950] [Medline: 22522112]

32. Nollen NL, Mayo MS, Carlson SE, Rapoff MA, Goggin KJ, Ellerbeck EF. Mobile technology for obesity prevention: a randomized pilot study in racial- and ethnic-minority girls. Am J Prev Med 2014 Apr;46(4):404-408 [FREE Full text] [doi: 10.1016/j.amepre.2013.12.011] [Medline: 24650843]

33. Blackman KC, Zoellner J, Berrey LM, Alexander R, Fanning J, Hill JL, et al. Assessing the internal and external validity of mobile health physical activity promotion interventions: a systematic literature review using the RE-AIM framework. J Med Internet Res 2013;15(10):e224 [FREE Full text] [doi: 10.2196/jmir.2745] [Medline: 24095951]

34. Fanning J, Mullen SP, McAuley E. Increasing physical activity with mobile devices: a meta-analysis. J Med Internet Res 2012;14(6):e161 [FREE Full text] [doi: 10.2196/jmir.2171] [Medline: 23171838]

35. Citrix ByteMobile. Mobile Analytics Report. 2015 Feb. URL: https://www.citrix.com/content/dam/citrix/en_us/documents/products-solutions/citrix-mobile-analytics-report-february-2015.pdf?accessmode=direct [accessed 2015-04-21] [WebCite Cache ID 6Xxrau7XJ]

36. Fiordelli M, Diviani N, Schulz PJ. Mapping mHealth research: a decade of evolution. J Med Internet Res 2013;15(5):e95 [FREE Full text] [doi: 10.2196/jmir.2430] [Medline: 23697600]

37. Direito A, Dale LP, Shields E, Dobson R, Whittaker R, Maddison R. Do physical activity and dietary smartphone applications incorporate evidence-based behaviour change techniques? BMC Public Health 2014;14:646 [FREE Full text] [doi: 10.1186/1471-2458-14-646] [Medline: 24965805]

38. Direito A, Jiang Y, Whittaker R, Maddison R. Smartphone apps to improve fitness and increase physical activity among young people: protocol of the Apps for IMproving FITness (AIMFIT) randomized controlled trial. BMC Public Health 2015;15(1):635 [FREE Full text] [doi: 10.1186/s12889-015-1968-y] [Medline: 26159834]

39. Schulz KF, Altman DG, Moher D. CONSORT 2010 statement: updated guidelines for reporting parallel group randomized trials. Ann Intern Med 2010 Jun 1;152(11):726-732. [doi: 10.7326/0003-4819-152-11-201006010-00232] [Medline: 20335313]

40. Eysenbach G. CONSORT-EHEALTH: improving and standardizing evaluation reports of Web-based and mobile health interventions. J Med Internet Res 2011;13(4):e126 [FREE Full text] [doi: 10.2196/jmir.1923] [Medline: 22209829]

41. Prochaska JJ, Sallis JF, Long B. A physical activity screening measure for use with adolescents in primary care. Arch Pediatr Adolesc Med 2001 May;155(5):554-559. [Medline: 11343497]

42. Nader PR, Bradley RH, Houts RM, McRitchie SL, O'Brien M. Moderate-to-vigorous physical activity from ages 9 to 15 years. JAMA 2008 Jul 16;300(3):295-305. [doi: 10.1001/jama.300.3.295] [Medline: 18632544]

43. 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;12(1):e4 [FREE Full text] [doi: 10.2196/jmir.1376] [Medline: 20164043]

44. Michie S, Abraham C, Whittington C, McAteer J, Gupta S. Effective techniques in healthy eating and physical activity interventions: a meta-regression. Health Psychol 2009 Nov;28(6):690-701. [doi: 10.1037/a0016136] [Medline: 19916637]

45. Six to Start. Zombies, run! 5k training URL: https://www.zombiesrungame.com/ [accessed 2015-04-21] [WebCite Cache ID 6XxqLIrln]

46. Splendid Things. Get Running (Couch to 5k) URL: http://splendid-things.co.uk/getrunning/ [accessed 2015-04-21] [WebCite Cache ID 6XxqSPOZf]

47. Macpherson H. Pragmatic clinical trials. Complement Ther Med 2004;12(2-3):136-140. [doi: 10.1016/j.ctim.2004.07.043] [Medline: 15561524]

48. The Cooper Institute. In: Meredith M, Welk G, editors. Fitnessgram & Activitygram Test Administration Manual-Updated 4th Edition. Dallas, TX: Human Kinetics; 2013.

49. Cureton KJ, Sloniger MA, O'Bannon JP, Black DM, McCormack WP. A generalized equation for prediction of VO2peak from 1-mile run/walk performance. Med Sci Sports Exerc 1995 Mar;27(3):445-451. [Medline: 7752874]

50. de Onis M, Onyango AW, Borghi E, Siyam A, Nishida C, Siekmann J. Development of a WHO growth reference for school-aged children and adolescents. Bull World Health Organ 2007 Sep;85(9):660-667 [FREE Full text] [Medline: 18026621]

51. Janz KF, Lutuchy EM, Wenthe P, Levy SM. Measuring activity in children and adolescents using self-report: PAQ-C and PAQ-A. Med Sci Sports Exerc 2008 Apr;40(4):767-772. [doi: 10.1249/MSS.0b013e3181620ed1] [Medline: 18317366]

52. Kowalski K, Crocker P, Kowalski N. Convergent validity of the physical activity questionnaire for adolescents. Pediatric Exercise Science 1997 Nov;9:342-352.

53. Dunton GF, Tscherne J, Rodriguez D. Factorial validity and gender invariance of the physical activity enjoyment scale (PACES) in older adolescents. Res Q Exerc Sport 2009 Mar;80(1):117-121. [doi: 10.1080/02701367.2009.10599543] [Medline: 19408473]

54. Motl RW, Dishman RK, Saunders R, Dowda M, Felton G, Pate RR. Measuring enjoyment of physical activity in adolescent girls. Am J Prev Med 2001 Aug;21(2):110-117. [Medline: 11457630]

55. Wilson P, Rogers W, Rodgers W, Wild T. The psychological need satisfaction in exercise scale. Journal of Sport & Exercise Psychology 2006;28:231-251.

56. Motl RW, Dishman RK, Trost SG, Saunders RP, Dowda M, Felton G, et al. Factorial validity and invariance of questionnaires measuring social-cognitive determinants of physical activity among adolescent girls. Prev Med 2000 Nov;31(5):584-594. [doi: 10.1006/pmed.2000.0735] [Medline: 11071840]

57. Bartholomew J, Loukas A, Jowers E, Allua S. Validation of the physical activity self-efficacy scale: testing measurement invariance between Hispanic and Caucasian children. J Phys Act Health 2006;3:70-78.

58. Mâsse LC, Fuemmeler BF, Anderson CB, Matthews CE, Trost SG, Catellier DJ, et al. Accelerometer data reduction: a comparison of four reduction algorithms on select outcome variables. Med Sci Sports Exerc 2005 Nov;37(11 Suppl):S544-S554. [Medline: 16294117]

59. Basterfield L, Adamson AJ, Pearce MS, Reilly JJ. Stability of habitual physical activity and sedentary behavior monitoring by accelerometry in 6- to 8-year-olds. J Phys Act Health 2011 May;8(4):543-547. [Medline: 21597127]

60. Trost SG, Loprinzi PD, Moore R, Pfeiffer KA. Comparison of accelerometer cut points for predicting activity intensity in youth. Med Sci Sports Exerc 2011 Jul;43(7):1360-1368. [doi: 10.1249/MSS.0b013e318206476e] [Medline: 21131873]

61. Evenson KR, Catellier DJ, Gill K, Ondrak KS, McMurray RG. Calibration of two objective measures of physical activity for children. J Sports Sci 2008 Dec;26(14):1557-1565. [doi: 10.1080/02640410802334196] [Medline: 18949660]

XSL•FO
RenderX

62.   The President's Challenge Program. 2013. Physical fitness test scores-normative data spreadsheet (excel) URL: https:/
        /www.presidentschallenge.org/tools-resources/forms.shtml [accessed 2015-04-21] [WebCite Cache ID 6XxrPKEbZ]

63.   Baquet G, van Praagh E, Berthoin S. Endurance training and aerobic fitness in young people. Sports Med
        2003;33(15):1127-1143. [Medline: 14719981]

64.   Goh G, Tan NC, Malhotra R, Padmanabhan U, Barbier S, Allen JC, et al. Short-term trajectories of use of a caloric-monitoring
        mobile phone app among patients with type 2 diabetes mellitus in a primary care setting. J Med Internet Res 2015;17(2):e33
        [FREE Full text] [doi: 10.2196/jmir.3938] [Medline: 25648130]

65.   Helander E, Kaipainen K, Korhonen I, Wansink B. Factors related to sustained use of a free mobile app for dietary
        self-monitoring with photography and peer feedback: retrospective cohort study. J Med Internet Res 2014;16(4):e109
        [FREE Full text] [doi: 10.2196/jmir.3084] [Medline: 24735567]

66.   Brannon EE, Cushing CC. A systematic review: is there an app for that? Translational science of pediatric behavior change
        for physical activity and dietary interventions. J Pediatr Psychol 2015 May;40(4):373-384. [doi: 10.1093/jpepsy/jsu108]
        [Medline: 25502745]

67.   Ryan R, Frederick C, Lepes D, Rubio N, Sheldon K. Intrinsic motivation and exercise adherence. Int J Sport Psychol
        1997;28:335-354.

Abbreviations

BCT: behavior change technique
BMI: body mass index
CRF: cardiorespiratory fitness
MVPA: moderate-to-vigorous physical activity
PA: physical activity
PACES: Physical Activity Enjoyment Scale
PAQ-A: Physical Activity Questionnaire for Adolescents
PASES: Physical Activity Self-Efficacy Scale
PNSES: Psychological Need Satisfaction in Exercise Scale
RCT: randomized controlled trial
SMS: short message service

XSL•FO
RenderX

Original Paper

Assessment of Web-Based Consumer Reviews as a Resource for Drug Performance

Swarnaseetha Adusumalli[1], MSc; HueyTyng Lee[1], BSc (Hons); Qiangze Hoi[1], BSc (Hons); Si-Lin Koo[2], MBBS; Iain Beehuat Tan[1,2,3], MBBS, PhD; Pauline Crystal Ng[1], PhD

[1]Genome Institute of Singapore, Singapore, Singapore

[2]National Cancer Centre Singapore, Singapore, Singapore

[3]Duke-NUS Graduate Medical School, Singapore, Singapore

Corresponding Author:
Pauline Crystal Ng, PhD
Genome Institute of Singapore
60 Biopolis St.
Singapore,
Singapore
Phone: 65 6808 8310
Fax: 65 6808 8300
Email: ngpc4@gis.a-star.edu.sg

Abstract

Background: Some health websites provide a public forum for consumers to post ratings and reviews on drugs. Drug reviews are easily accessible and comprehensible, unlike clinical trials and published literature. Because the public increasingly uses the Internet as a source of medical information, it is important to know whether such information is reliable.

Objective: We aim to examine whether Web-based consumer drug ratings and reviews can be used as a resource to compare drug performance.

Methods: We analyzed 103,411 consumer-generated reviews on 615 drugs used to treat 249 disease conditions from the health website WebMD. Statistical analysis identified 427 drug pairs from 24 conditions for which two drugs treating the same condition had significantly and substantially different satisfaction ratings (with at least a half-point difference between Web-based ratings and $P<.01$). PubMed and Google Scholar were searched for publications that were assessed for concordance with findings online.

Results: Scientific literature was found for 77 out of the 427 drug pairs and compared to findings online. Nearly two-thirds (48/77, 62%) of the online drug trends with at least a half-point difference in online ratings were supported by published literature ($P=.02$). For a 1-point online rating difference, the concordance rate increased to 68% (15/22) ($P=.07$). The discrepancies between scientific literature and findings online were further examined to obtain more insights into the usability of Web-based consumer-generated reviews. We discovered that (1) drugs with FDA black box warnings or used off-label were rated poorly in Web-based reviews, (2) drugs with addictive properties were rated higher than their counterparts in Web-based reviews, and (3) second-line or alternative drugs were rated higher. In addition, Web-based ratings indicated drug delivery problems. If FDA black box warning labels are used to resolve disagreements between publications and online trends, the concordance rate increases to 71% (55/77) ($P<.001$) for a half-point rating difference and 82% (18/22) for a 1-point rating difference ($P=.002$). Our results suggest that Web-based reviews can be used to inform patients' drug choices, with certain caveats.

Conclusions: Web-based reviews can be viewed as an orthogonal source of information for consumers, physicians, and drug manufacturers to assess the performance of a drug. However, one should be cautious to rely solely on consumer reviews as ratings can be strongly influenced by the consumer experience.

Introduction

When choosing among drugs to treat a patient's condition, clinicians rely on published clinical trials, practice experience, and/or US Food and Drug Administration (FDA) drug labels. However, FDA trial results can be incomplete; 78% of drug trials subject to mandatory reporting did not report their results [1]. Furthermore, published results may be reported in a biased manner, favoring the trial sponsor, which is often also the drug manufacturer [2,3]. Finally, published trials and FDA labels can be challenging to read and inaccessible to patients.

The public is increasingly turning to the Internet for information about drugs and their side effects [4,5]. Three-quarters (73%) of adults with higher education use the Internet for health information [4]. Over a quarter (26%) of Americans read or watch someone else's experience with health or medical issues online [5], and 16% of Internet users go online to find others who share the same health concerns [5].

Many of the health-related websites, such as WebMD [6] and AskAPatient [7], provide a public forum for consumers to post ratings and reviews on their drug experiences. For example, consumers can submit their reviews of drugs and rate the drugs on a scale of 1-5 at the WebMD website. The websites also ask users to share their disease condition, age, sex, the prescribed duration, and their comments; 3-4% of Internet users have shared their experiences with drugs online [5], which extrapolates to millions of drug experiences.

Various researchers have mined data from health websites to cull useful information from users' comments [8-14]. Past research on health-related websites and communities has focused on text mining reviewers' comments. One research group collected consumer reviews of statins from three health-related websites and were able to associate statin consumption with side effects that were not listed by the drug manufacturer but supported by other studies [10]. In another study, consumer reviews and professional drug descriptions reported similar efficacies and adverse effects for two psychotropic drugs [11]. A study on Parkinson's disease showed that online forums may be a useful source of observational information to complement clinical trials [14]. Past studies have successfully compared drug performance with online resources, but on a case-by-case basis, by focusing on certain classes of drugs or drugs treating a single condition [10-13].

In this study, we investigate if Web-based review ratings can be used as a resource to compare drug performance on a global scale for a comprehensive set of drugs treating a variety of disease conditions. Web-based review ratings potentially provide a fast and easily accessible data source for drugs. We sought to determine if crowd-sourced review ratings are supported by published literature and if they can provide a complementary resource to clinical trials.

Methods

Drug Comparison Based on Web-Based Reviews

Consumer reviews are publicly available and anonymous, so it is ethically acceptable to conduct an analysis of the comments without seeking informed consent from their authors [15]. We obtained an exemption from the *Institutional Review Board* to analyze online consumer-generated reviews.

We downloaded 141,210 reviews of 1503 drugs treating 1123 conditions from WebMD on October 23, 2012. Drug and condition names were taken from the WebMD website. Each review had a user satisfaction rating. The satisfaction rating ranged from 1-5, where 1 is the lowest score for expressing dissatisfaction and 5 is the highest score for expressing satisfaction with the drug. In addition to these ratings, we downloaded the genders and ages of the reviewers and the text comments of the reviews.

We applied pre-processing steps prior to statistical analysis. First, drugs with different modes of deliveries for each individual condition were grouped separately (eg, oral versus intravenous). Second, the reviews of drugs with the same active ingredient(s) were combined. Information about drugs' brand names and active ingredients was downloaded from the Drugs@FDA database [16]. Of the 1503 drugs on the WebMD drug list, the active ingredients of 920 (61.21%) drugs were identified. For drugs whose active ingredient was not listed in the FDA drug database, the original drug name was kept in the subsequent analysis. Thus, the 1503 drugs were reduced to a total of 1215 groups of active ingredients/drugs, which were used for the analysis. To be concise, we refer to an active ingredient group as a drug in this study. Drugs were required to have at least 30 reviews for the particular condition, and the condition was required to have at least two drug groups to be selected for analysis. This gave a final list of 249 conditions encompassing 615 drugs and 103,411 reviews.

We first tested whether drug ratings were significantly different within a disease condition, before examining drugs individually. We tested at the level of disease condition for two reasons: (1) to control for patient heterogeneity as much as possible, with the assumption that patients taking drugs for the same condition would have similar patient profiles, and (2) because testing for all pairwise drug combinations across all conditions would require a large Bonferroni correction factor, whereas testing for conditions bounds the correction factor to the smaller number of conditions (n=249). Analysis of covariance (ANCOVA) was applied to each condition to determine whether drug(s) account for significant differences in satisfaction ratings while controlling for the covariates of gender and age. For each condition, a linear model was constructed with drug, age, and gender as independent variables, and the satisfaction rating as the dependent variable. Age ranges were transformed into numeric values by taking the mean of the age range (eg, a reviewer with an age range of 25-34 years was assigned 29.5). We computed ANCOVA for the linear model using the "car" library in R (ANOVA, type="III"). We identified 24 conditions that had a statistically significant difference between their drugs' ratings ($P<.05$, after Bonferroni correction for the 249 conditions tested).

For each of the 24 conditions, we focused on comparing drugs with significant and substantially different ratings. Because comparing two drugs with minor rating differences is difficult, we examined drug pairs where the two drugs' adjusted drug

ratings differed by at least 0.5 points. Adjusted drug ratings are controlled for gender and age because the two drugs may have slightly different patient distributions. An adjusted drug rating is computed by taking the predicted value of the drug's score for the most common age and gender for the condition (using R's predict function for the linear model). Additionally, the two drugs were required to have significantly different online satisfaction ratings (Mann-Whitney U test, $P<.01$). In summary, 427 drug pairs with significantly and substantially different ratings were identified from the 24 conditions (see Figure 1 for an example and Multimedia Appendix 1 for the full table). For example, felodipine has an adjusted online rating of 3.2, which is 0.7 points higher than amlodipine (adjusted online rating 2.5), and the two drugs' online ratings are significantly different from each other ($P=.003$). Online trends can be deduced from each drug pair. In the aforementioned example, the deduced online trend is that felodipine is a better drug than amlodipine. Thus, 427 online trends are deduced from the 427 drug pairs.
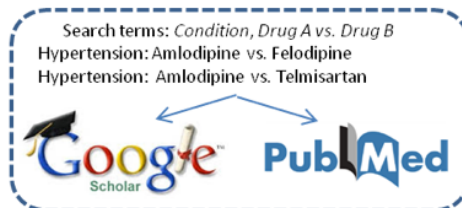
Figure 1. Examples of drug pairs with significant and substantially different ratings, and a procedure flowchart for comparing online findings with scientific literature.

Comparison With Literature

Overview

The aim of this study is to see whether scientific literature supports the deduced online trends. This was done by mining the literature for a comparison of the two drugs from that particular online trend. Figure 1 shows an overview of the steps, using three drugs that treat hypertension as an example.

Step 1. Search the Scientific Literature

Literature searches were carried out for all 427 drug pairs with significantly and substantially different Web-based ratings. Because the WebMD condition's name may not be standard, MedDRA's preferred term for the condition name was used for the literature search [17] (see Multimedia Appendix 2 for condition mappings). The literature searches were conducted on Google Scholar and PubMed. Each search was limited to three publications for practicality, so the maximum number of publications each pairwise drug comparison could have is six (three for Google Scholar and three for PubMed). Occasionally, official regulatory bodies such as the FDA were also used as a source because their documents would be discovered by Google Scholar (eg, see Armour Thyroid vs levothyroxine; Multimedia Appendix 3).

Publications were required to have the drug name exactly and treat the same condition as the WebMD listing. We discarded publications that did not pertain to humans (eg, studies in rats) and case reports on single patients. For 82 out of the 427 pairwise comparisons, we found 152 pieces of scientific literature (132 head-to-head comparisons, 11 reviews, three meta-analyses, and six others).

Step 2. Assess the Better Drug in the Publication

The better performing drug was interpreted from a publication's abstract. Two authors read the abstract and decided whether one drug performed better than the other, if the two drugs performed similarly, or if performance was unclear. An example of an unclear performance is if drug A is more effective but has worse side effects than drug B. If two authors disagreed on the classification, the abstract was discussed between the 2 authors until an agreement was reached. The decision-making process for head-to-head comparisons, meta-analyses, regulatory bodies, and review articles was identical.

Step 3. Compare a Publication's Conclusion With the Corresponding Online Trend

A verdict was determined as to whether the better drug from a publication concurred with the better drug from the corresponding online trend. We classified each publication as "agree" when the publication's abstract agreed with the online trend and "disagree" when the publication disagreed with the online trend (see Figure 1). An example of an "agree" publication is a paper that states: "Telmisartan was more effective than amlodipine in preventing AF [atrial fibrillation] recurrences" [18]. According to WebMD, telmisartan has a higher Web-based rating than amlodipine (3.1 vs 2.5). Therefore, the publication agrees with the online trend that telmisartan is the better drug. A "disagree" verdict is given when the publication states that the poorly rated WebMD drug has better

or similar performance, or if the publication's conclusion was unclear.

Step 4. For Each Deduced Online Trend, Combine the Publications Into a Single Verdict

Because a single online trend can have multiple publications, the publications' agree/disagree statuses are summarized into a single verdict. The verdict was concluded as "agree" when the majority of the publications for that comparison agreed with the deduced online trend. The verdict was concluded to "disagree" when the majority of publications disagreed with the deduced online trend. For example, if a drug comparison had four published studies, of which three agreed with the deduced online trend and one did not, we concluded the verdict to "agree" with the deduced online trend. For five pairwise drug comparisons, an equal number of publications agreed and disagreed with WebMD ratings; these inconclusive drug comparisons were removed from consideration. In total, there were 77 online trends with .50 point difference that had verdicts summarized from 141 pieces of scientific literature. For these 77 online trends, 48 and 29 had "agree" and "disagree" verdicts with the scientific literature, respectively (Multimedia Appendix 3). Most (71/77, 92%) of the verdicts were unanimous verdicts, where all the publications agreed with each other as to which was the better drug (Multimedia Appendix 3).

To determine if the observed number of "agree" verdicts was more than expected by chance, the P value for publication support was calculated by assuming a 0.50 probability of "agree" verdicts and a 0.50 probability of "disagree" verdicts. The probability of observing at least n number of "agree" verdicts was calculated using a cumulative binomial distribution.

Information Extraction From Food and Drug Administration Labels

FDA labels were used to reconcile the disagree verdicts between publications and deduced online trends. A drug's FDA label was used to determine the drug's serious side effects, off-label use, and addictive properties. To investigate serious side effects, we inspected a drug's FDA label for a black box warning, which is the strictest warning by the FDA. To see whether a drug is being used off-label, we looked at the conditions listed under the "Indications and usage" section. If the WebMD/MedDRA condition was not listed in the Indications section, we deemed this "off-label" use. To identify drugs with addictive properties, we inspected if a drug's FDA label noted drug abuse and dependence as a side effect.

The purpose of examining FDA labels was to find differences between drugs. If both of the drugs in the pairwise comparison had black box warnings or both drugs had addictive properties, this was not recorded as an observation because the two drugs were similar for that aspect.

Text Mining of Reviews

For some drugs, we examined the reviewers' comments to hypothesize why publications and deduced online trends might disagree. Frequencies of certain words were counted in the comment section of reviews. The number of drug reviews that contained the term was divided by the total number of drug

XSL•FO

RenderX

reviews. Statistical significance for difference in word frequencies was calculated using the chi-square test.

For type 2 diabetes, reviewers' comments were searched for the word "heart" because the poorly rated drug pioglitazone had a black box warning for congestive heart failure. When looking at addictive drugs (carisoprodol, nefazodone hydrochloride, and diazepam), reviewers' comments were searched for the words "abuse" and "addict".

For asthma, we found the most frequent words among reviewers' comments by using word clouds. Reviewers' comments were fed to Voyant Tools [19], which removes stop words from the reviews and generates a word cloud from the remaining words. For the drug ProAir, "inhaler" was the most frequent word. For albuterol, ProAir's generic equivalent, "asthma" was the most frequent word.

Results

### Drug Differences Identified From Web-Based Reviews

Our study investigates the usefulness of Web-based rating differences between drugs. Over 140,000 drug reviews were downloaded from WebMD. To detect drug rating differences, ANCOVA analysis was applied to 249 disease conditions, of which 24 had different performances between drugs (see Methods). Within the 24 conditions, there were 427 drug pairs that had substantially and significantly different ratings, with at least .50 point difference between the two drugs' ratings (P<.01 Mann-Whitney) (see Multimedia Appendix 1 and Methods).

For each drug pair, one can deduce an online trend because one drug rates significantly higher than the other drug. For example, felodipine has a higher online rating than amlodipine (3.2 vs 2.5, P=.003), so the deduced online trend is that felodipine is a better drug than amlodipine. Examples for two pairwise drug comparisons pertaining to the condition hypertension are found in Figure 1.

To assess if deduced online trends were concordant with scientific literature, we manually searched PubMed and Google Scholar for publications that compare the two drugs belonging to the online trend (see Methods and Figure 1). A verdict was determined as to whether the majority of the publications agreed with the online drug trend. Verdicts were determined for 77 (18.0%) of the 427 drug pairs (see examples in Figure 1 and the full table in Multimedia Appendix 3). Summarizing across the 77 pairwise drug comparisons with at least a half-point rating difference, 62% (48/77) of the literature verdicts are concordant with their deduced online trends (P=.02, binomial distribution) (Figure 2). When raising the cutoff to a 1-point rating difference between two drugs, the concordance rate between Web-based ratings and literature increases to 68% (15/22). The result for the higher cutoff is not significant, most likely due to low numbers (P=.07) (Figure 2).

Figure 2. Concordance between deduced online trends and scientific support at varying levels of point differences between 2 drugs' online ratings. The solid line indicates the concordance of online trends with literature and the dashed line indicates the concordance of online trends with scientific literature and FDA labels. For each data point, the percentage concordance is shown and the number of drug pairwise comparisons agreeing with scientific support divided by the total number of drug pairwise comparisons are given in parenthesis. The asterisk indicates statistical significance with $P<.05$ according to the binomial test.
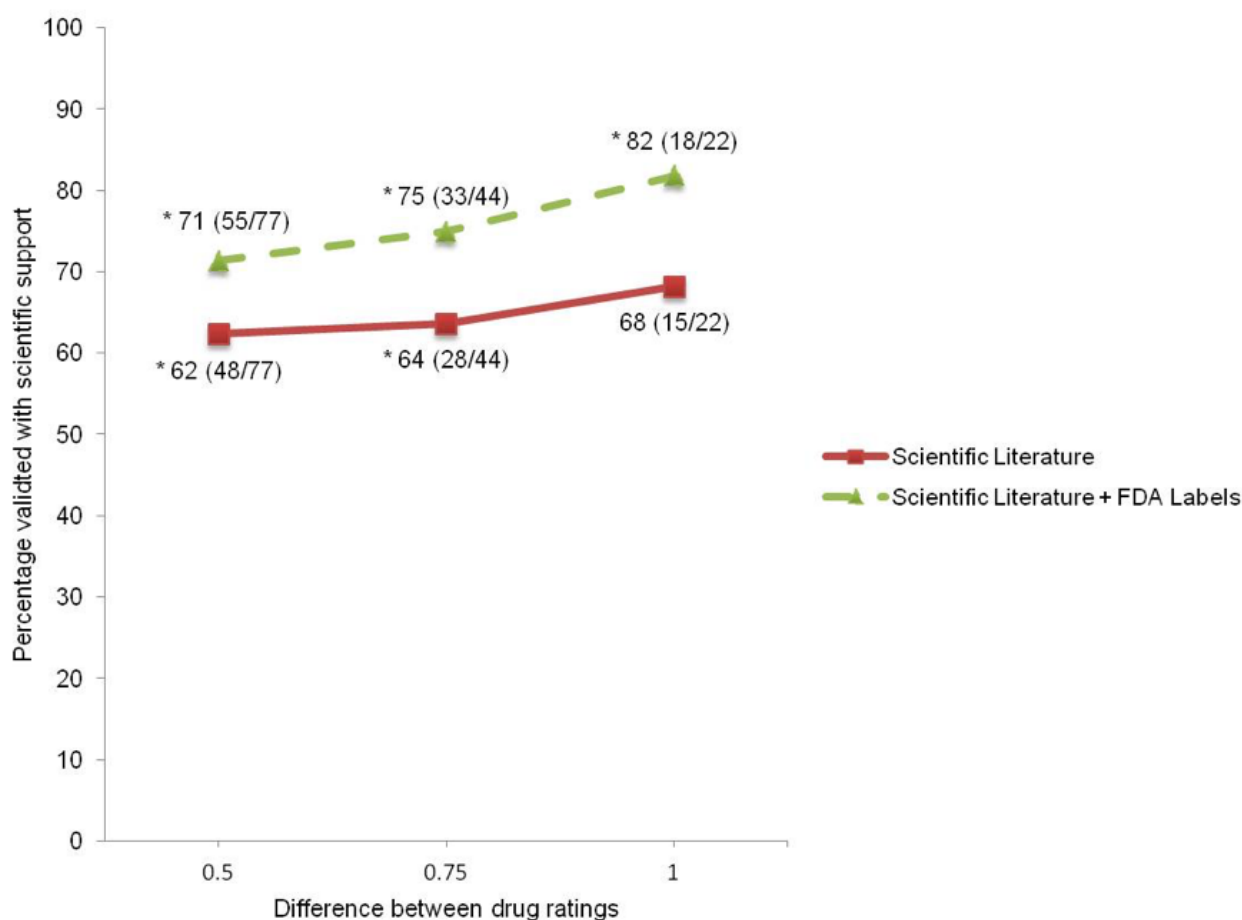
Investigating Online Trends That Are Discordant With Scientific Literature

While the majority of deduced online trends were in concordance with the literature, 38% (29/77) were not. We investigated why scientific literature was not consistent with Web-based ratings. We observed that (1) drugs with FDA boxed warnings or used off-label for the WebMD condition rated poorly among online reviews, (2) drugs with addictive properties had higher review ratings, and (3) patients rated alternative treatments higher. A problem with drug delivery was also discovered independently. The summary of these findings can be found in Table 1 (with additional details in Multimedia Appendix 4) and is further elaborated on in the following text.

Table 1. Summary of observations for drug comparisons where Web-based ratings disagreed with publications.

	Number of drug comparisons
Consistent with FDA label	
Drug with boxed warning rated lower	7
Drug used off-label rated lower	2
Addictive drug rated higher	5
Alternative or second-line drug rated higher	2
Unexplained	13
Total	29

Poorly Rated Drugs Have Food and Drug Administration Black Box Warnings

FDA drug labels can have black box warnings that inform of serious side effects. For seven drug comparisons, the drugs with FDA black box warnings were poorly rated among Web-based reviews even though they performed better according to publications (Multimedia Appendix 4). The corresponding competing drugs in the pairwise comparison did not have black box warnings. For example, the type 2 diabetes drug pioglitazone hydrochloride has a black box warning of increased risk of congestive heart failure [20]. Approximately 9.9% (56/568) of pioglitazone hydrochloride reviewers complained of heart problems, compared with 3.08% (55/1788) of the complaints from other drugs used to treat type 2 diabetes (chi-square $P<.001$) (see Methods). Sample comments for pioglitazone hydrochloride are "it is a killer!!! better choices are available. who needs bladder cancer or heart problems!!!" and "had very bad chest pains thought i was having heart failure. chest pains stop[p]ed within 3 days of stop[p]ing the drug will not try it again. it did lower blood sugars". Thus poor Web-based ratings for certain drugs are supported by FDA black box warnings.

If one assumes FDA black box warnings are accurate and authoritative compared to scientific publications (which can be biased [2,3]), then the seven discordant comparisons may be considered correct for the deduced online trends. If we re-classify these seven to a verdict of "agreed", then the support rate of online ratings is 71% ((48+7)/77) ($P<.001$) (see dashed line in Figure 2). This is an overestimate of the support rate because we investigated only pairwise comparisons with "disagree" verdicts. Future studies should include a more thorough analysis examining FDA labels for all drugs.

Off-Label Drug Use Is Rated Lower

Drugs are sometimes used to treat a condition that has not been approved by the FDA. The practice of off-label drug use is prevalent [21]. For two comparisons, the drug with a lower Web-based rating was not FDA-approved for the WebMD condition (Multimedia Appendix 4). For example, alprazolam has the highest rating compared to all other drugs for panic disorder and is one of the few drugs indicated for panic disorder on its FDA label. Most of the other drugs that have reviews on WebMD for panic disorder are indicated to treat depression (eg, citalopram) or anxiety disorder (eg, diazepam), and these other drugs have lower Web-based ratings. Therefore, even though the practice of off-label drug prescription is common, Web-based reviews can reveal user dissatisfaction with off-label drugs.

Addictive Drugs Are Highly Rated

Three drugs (diazepam for treating anxiety and muscle spasms, nefazodone hydrochloride for treating depression, and carisoprodol for treating muscle spasms) are addictive according to FDA labels. These drugs have poor performances according to the scientific literature, but higher Web-based ratings compared to other drugs treating the same condition. This suggests the possibility that patients may rate drugs with addictive properties higher.

For example, carisoprodol (adjusted rating 4.23) is rated higher than the other drugs that treat muscle spasm (Multimedia Appendix 1). Head-to-head comparisons showed that carisoprodol and cyclobenzaprine (adjusted rating 3.26) perform similarly, but carisoprodol's usefulness is mitigated by its potential for abuse [22]. The drug has an FDA label that warns of its addictive properties. Interestingly, 12.6% (28/223) of the reviewers' comments for carisoprodol contained the word "addict" or "abuse", compared with 0% (0/371) for cyclobenzaprine reviews ($P<.001$) (see Methods). However, even though carisoprodol reviewers recognized the potential for abuse, they rated carisoprodol highly: 68% (19/28) of the reviewers that mentioned addictiveness gave a satisfaction score of 4 or higher. Some sample comments are "It's a great medication but can easily become dangerously addicting" and "so far this has been the best medication to help give me almost complete relief. Just be careful using it, it is addictive."

Similarly, for the addictive drug diazepam, 87% (13/15) of the reviewers for anxiety and muscle spasm that mentioned "addict" or "abuse" still gave ratings 4 or higher. This suggests that patients, despite being aware of a drug's potential for abuse, will still rate an addictive drug high. It highlights the importance of professional medical advice and FDA labels, and a caution when relying on consumer-generated reviews. Another possible explanation for why drugs with addictive properties are rated higher may be due to stronger drug efficacy and potency or psychoactive properties. A more systematic study of the impact on addictiveness in Web-based ratings should be conducted to see if these observations can be generalized.

Alternative Treatments Can Have Higher Ratings

Drug accessibility and past experience may influence reviewers' drug ratings. There were two drug comparisons for which an alternative or second-line drug was rated higher than the commonly prescribed first-line drug (Multimedia Appendix 4). For hypothyroidism, online patients expressed greater satisfaction with the drug Armour (adjusted rating 3.92) than levothyroxine (adjusted rating 2.22) (Multimedia Appendix 1). Treatment with Armour is highly controversial. Armour is desiccated animal thyroid, and this natural treatment has been used to treat hypothyroidism since the 1890s [23], prior to the formation of the FDA. Levothyroxine is a synthetic form of the thyroid hormone and is FDA-approved [24-26]. Despite professional endorsements of the synthetic form because of its better stability and quality assurance (by United States FDA, Endocrine Society of Australia, and British Thyroid Association), there is a grassroots movement in support of using naturally desiccated thyroid [27]. Head-to-head comparisons exist, but interpretations of these comparisons are controversial [28]. In surveys, patients preferred natural desiccated thyroid over thyroxine alone [28], and the majority of patients who had tried conventional therapies but then switched to natural desiccated thyroid were more satisfied with the natural treatment [29]. Web-based patient reviews are consistent with surveys rather than the professional recommendations, and the higher rating may be due to subgroups that are satisfied with Armour as an alternative treatment.

These results suggest ratings can be influenced by a reviewer's treatment history. If the first line of treatment is ineffective and the alternative treatment provides relief but is harder to obtain, reviewers may compensate with higher ratings for the alternative/second-line drug to confirm that the less popular or less common choice was effective for them. A more systematic study is necessary to see if this trend can be generalized.

In summary, Web-based ratings that disagree with scientific literature can be explained by (1) drugs with FDA boxed warnings rating poorly, (2) drugs used for off-label conditions rating poorly, (3) drugs with addictive properties rating higher, and (4) alternative treatments rating higher. These explanations account for over half (16/29) of the discordances between literature and deduced online trends (Table 1). The remaining 13 disagreements were designated as "Unexplained". Further investigation is needed to reconcile the remaining cases.

Drug Delivery Design

Web-based reviews can lead to new findings; a drug delivery issue for an asthma inhaler was discovered. This came to our attention because the asthma inhaler ProAir had low Web-based ratings (average rating 1.46), yet its generic equivalent albuterol had high Web-based ratings (average rating 3.48). We observed this strange phenomenon when we had not yet combined the brand-name ProAir with its generic equivalent albuterol. To understand this unexpected discrepancy, we inspected the text of the reviews. The most frequent word in the ProAir reviews is "inhaler", suggesting that dissatisfaction with ProAir was due to the inhaler's design. Some comments on the inhaler include: "This inhaler continually clogs and I waste quite a bit of medication" and "ProAir frequently clogs and never really seems to dispense properly. Its effectiveness is a large step backwards from fast acting inhalers 10 years ago".

The company responded by releasing a newly designed inhaler in 2012, which included a dosage counter capable of tracking the number of doses remaining in the inhaler [30]. Findings related to drug-delivery issues may not be assessed in clinical trials. Therefore, consumer input from Web-based reviews can extend beyond the efficacy of the active ingredient and can benefit the drug manufacturer.

Discussion

Principal Findings

Previous publications have studied drug reviews using online resources, but these approaches tend to examine drugs on a case-by-case basis [10-13]. To the best of our knowledge, this is the first study to analyze online drug satisfaction on a global scale for a comprehensive set of drugs across many disease conditions. We found 427 significantly different drug pairs where the drugs' ratings had more than a half-point difference. For 77 of the drug pairs, we determined whether the scientific literature agreed or disagreed with the deduced online trends. For a 0.5-point rating difference, 62% (48/77) of the deduced online trends were concordant with scientific literature (P=.02). The concordance increased to 68% (15/22) when drug pairs with a larger rating difference (at least 1-point) were considered, but this was not statistically significant (P=.07), possibly due

to small sample size. Further investigation of the remaining 29 that were discordant showed that seven inferred online findings were supported by FDA labels. Lower-rated drugs had FDA black box warnings indicating serious side effects. If one were to include the FDA black box warnings as supportive evidence for the deduced online trends, the scientific support for online trends increases to 71% (55/77) (P<.001).

Examination of the discordant drug comparisons suggested reviewers may be rating addictive drugs and alternative drugs higher. Addictive and alternative drugs may have similar efficacy to non-addictive and standard drugs; high ratings could be an artifact of users' subjectivity. These observations were found for a small number of drug comparisons and may be anecdotal. A more comprehensive study is necessary before generalizing if addictive drugs or second-line drugs tend to have higher ratings.

Web-based reviews also uncovered a new finding: the suboptimal design of an asthma inhaler. Such analyses can assess the satisfaction of a drug beyond the efficacy of its active ingredient as features like drug delivery may not always be assessed in clinical trials. A drug manufacturer can use this knowledge to improve the delivery design and manufacturing process.

Limitations

The use of Web-based reviews is independent, fast, and inexpensive, but it also poses some challenges. The reviewers themselves may be biased. People who write reviews may be different from the general population. Reviewers provide a subjective rating on "satisfaction" and do not have objective criteria to assess clinical benefit, unlike the "harder" endpoints that are evaluated in clinical studies. Users experience a drug's effects on a broader spectrum than the narrowly defined efficacy endpoints of clinical drug studies. This could cause the differences between quantitative Web-based ratings and published drug efficacies. Our study also suggests that reviewers may downplay certain side effects, such as addictiveness. Another disadvantage is that most review websites do not require information on important clinical input variables such as dosages, drug compliance, duration of treatment, additional drugs taken, strict diagnostic criteria, uniform disease severity/stage, smoking status, and general health. Therefore, one cannot ensure that the patients receiving drug A have similar medical profiles to those receiving drug B. While an analysis based on consumer reviews may involve a certain degree of bias and caveats, it also measures the exposure of a drug in a more realistic and diverse setting. Another limitation of our study is that we used only one source for reviews; future work will be incorporation of other additional online sources.

Conclusion

A small number (3-4%) of Internet users have shared their experiences with drugs online [5], which extrapolates to millions of drug experiences. The large size and broad accessibility of this database has an advantage over controlled clinical trials that recruit a finite number of patients. This is counterbalanced by the fact that Internet users represent a diverse population, in contrast to controlled clinical trials, which consist of

homogeneous patients who meet strict trial inclusion criteria. Nevertheless, our study characterizes the use of Web-based reviews for comparing performances of drugs. In conclusion, we have shown that consumer reviews can be used as an orthogonal source to reveal insights on drug performance.

Acknowledgments

Conflicts of Interest

None declared.

Multimedia Appendix 1

Literature searches for 427 drug pairs where the two drugs have substantially different online ratings that are significantly different. The adjusted online drug ratings are shown in columns 4 and 5, and P values for Mann-Whitney test are in column 6.

[XLSX File (Microsoft Excel File), 34KB - jmir_v17i8e211_app1.xlsx]

Multimedia Appendix 2

Search terms for the 24 conditions with drug differences.

[PDF File (Adobe PDF File), 5KB - jmir_v17i8e211_app2.pdf]

Multimedia Appendix 3

Drug pairs for which scientific literature was found, and their verdicts as to whether the scientific literature agreed or disagreed with online trends. The second column shows the two drugs being compared with their adjusted online ratings in parentheses. The third column shows the scientific literature that compares the two drugs, and whether it disagrees or agrees with the deduced online finding in parentheses. The fourth column lists the number of scientific studies that agree with the deduced online finding out of the total number of relevant scientific studies. An overall verdict summarizing the multiple scientific studies is shown in the fifth column. The sixth column specifies whether the verdict is unanimous. The last column lists the publication type for each piece of scientific literature.

[XLSX File (Microsoft Excel File), 19KB - jmir_v17i8e211_app3.xlsx]

Multimedia Appendix 4

Possible explanations for cases where scientific literature disagrees with deduced online findings.

[PDF File (Adobe PDF File), 5KB - jmir_v17i8e211_app4.pdf]

References

1. Prayle AP, Hurley MN, Smyth AR. Compliance with mandatory reporting of clinical trial results on ClinicalTrials.gov: cross sectional study. BMJ 2012;344:d7373 [FREE Full text] [doi: 10.1136/bmj.d7373] [Medline: 22214756]
2. Bero L, Oostvogel F, Bacchetti P, Lee K. Factors associated with findings of published trials of drug-drug comparisons: why some statins appear more efficacious than others. PLoS Med 2007 Jun;4(6):e184 [FREE Full text] [doi: 10.1371/journal.pmed.0040184] [Medline: 17550302]
3. Krauth D, Anglemyer A, Philipps R, Bero L. Nonindustry-sponsored preclinical studies on statins yield greater efficacy estimates than industry-sponsored studies: a meta-analysis. PLoS Biol 2014 Jan;12(1):e1001770 [FREE Full text] [doi: 10.1371/journal.pbio.1001770] [Medline: 24465178]
4. Centers for Disease Control and Prevention (CDC). 2009. Use of the Internet for Health Information: United States URL: http://www.cdc.gov/nchs/data/databriefs/db66.htm [accessed 2014-10-24] [WebCite Cache ID 6TYbG9QqL]
5. Pew Research Center. 2013. Pew Internet & American Life Project Health Online URL: http://www.pewinternet.org/2013/01/15/health-online-2013/ [accessed 2014-10-24] [WebCite Cache ID 6TYYGPNd6]
6. WebMD. URL: http://www.webmd.com/ [accessed 2015-03-05] [WebCite Cache ID 6WnCfbOPT]
7. AskAPatient. URL: http://www.askapatient.com/ [accessed 2015-03-05] [WebCite Cache ID 6WnCq0qhl]
8. Chee BW, Berlin R, Schatz B. Predicting adverse drug events from personal health messages. AMIA Annu Symp Proc 2011;2011:217-226 [FREE Full text] [Medline: 22195073]

XSL•FO
RenderX

9. Leaman RW, Sullivan R, Skariah A, Yang J. Towards internet-age pharmacovigilance: extracting adverse drug reactions from user posts to health-related social networks. 2010 Presented at: Workshop on Biomedical Natural Language Processing; 2010; Stroudsburg, PA p. 117-125.

10. Liu J, Li A, Seneff S. Automatic Drug Side Effect Discovery from Online Patient-Submitted Reviews: Focus on Statin Drugs. 2011 Presented at: First International Conference on Advances in Information Mining and Management; 2011; Cambridge, MA URL: http://research.microsoft.com/apps/pubs/default.aspx?id=219851

11. Moncrieff J, Cohen D, Mason JP. The subjective experience of taking antipsychotic medication: a content analysis of Internet data. Acta Psychiatr Scand 2009 Aug;120(2):102-111. [doi: 10.1111/j.1600-0447.2009.01356.x] [Medline: 19222405]

12. Hughes S, Cohen D. Can online consumers contribute to drug knowledge? A mixed-methods comparison of consumer-generated and professionally controlled psychotropic medication information on the internet. J Med Internet Res 2011;13(3):e53 [FREE Full text] [doi: 10.2196/jmir.1716] [Medline: 21807607]

13. Abou TM, Rossard C, Cantaloube L, Bouscaren N, Roche G, Pochard L, et al. Analysis of patients' narratives posted on social media websites on benfluorex's (Mediator®) withdrawal in France. J Clin Pharm Ther 2014 Feb;39(1):53-55. [doi: 10.1111/jcpt.12103] [Medline: 24304185]

14. Schröder S, Zöllner YF, Schaefer M. Drug-related problems with Antiparkinsonian agents: consumer Internet reports versus published data. Pharmacoepidemiol Drug Saf 2007 Oct;16(10):1161-1166. [doi: 10.1002/pds.1415] [Medline: 17486665]

15. Eysenbach G, Till JE. Ethical issues in qualitative research on internet communities. BMJ 2001 Nov 10;323(7321):1103-1105 [FREE Full text] [Medline: 11701577]

16. The US Food and Drug Administration. URL: http://www.fda.gov/Drugs/InformationOnDrugs/ucm135821.htm [accessed 2014-10-24] [WebCite Cache ID 6TYdR12xv]

17. MedDRA. URL: http://www.meddra.org/ [accessed 2015-06-29] [WebCite Cache ID 6Zdz5ZjDs]

18. Fogari R, Zoppi A, Maffioli P, Mugellini A, Preti P, Perrone T, et al. Effect of telmisartan on paroxysmal atrial fibrillation recurrence in hypertensive patients with normal or increased left atrial size. Clin Cardiol 2012 Jun;35(6):359-364. [doi: 10.1002/clc.21994] [Medline: 22522403]

19. Sinclair S, Rockwell G, Voyant Tools Team. Voyant Tools (web application). 2012. URL: http://voyant-tools.org/ [accessed 2015-06-29] [WebCite Cache ID 6ZdxEseNq]

20. DailyMed. ACTOS- pioglitazone hydrochloride tablet URL: http://dailymed.nlm.nih.gov/dailymed/drugInfo.cfm?setid=d2ddc491-88a9-4063-9150-443b4fa4330c [accessed 2015-03-08] [WebCite Cache ID 6WsRmcydJ]

21. Walton SM, Schumock GT. Should off-label drugs be included as comparators in pharmacoeconomic studies? Pharmacoeconomics 2014 Nov;32(11):1035-1037. [doi: 10.1007/s40273-014-0222-2] [Medline: 25270597]

22. Toth PP, Urtis J. Commonly used muscle relaxant therapies for acute low back pain: a review of carisoprodol, cyclobenzaprine hydrochloride, and metaxalone. Clin Ther 2004 Sep;26(9):1355-1367. [doi: 10.1016/j.clinthera.2004.09.008] [Medline: 15530999]

23. Lindholm J, Laurberg P. Hypothyroidism and thyroid substitution: historical aspects. J Thyroid Res 2011:- [FREE Full text] [doi: 10.4061/2011/809341] [Medline: 21760981]

24. British Thyroid Association. 2007. Armour Thyroid (USP) and combined thyroxine/triodothyronine as thryoid hormone replacement URL: http://www.british-thyroid-association.org/info-for-patients/Docs/bta_Armour_T4_T3.pdf [accessed 2014-10-24] [WebCite Cache ID 6TYe2Wikl]

25. Ebeling PR. ESA Medical Affairs, ESA Position Statement on Dessicated Thyroid or Thyroid Extract. URL: https://endocrinesociety.org.au/ESA_Position_Statement_%20Thyroid-Extract.pdf [accessed 2015-03-05] [WebCite Cache ID 6Wnd59Am9]

26. Parks M. FDA Consumer Health Information: U.S. Food and Drug Administration. 2013. Thyroid Medications: Q & A with Mary Parks, M.D URL: http://www.fda.gov/ForConsumers/ConsumerUpdates/ucm107377.htm [accessed 2014-10-24] [WebCite Cache ID 6TYeGg968]

27. Bowthorpe JA. Stop The Thyroid Madness. URL: http://www.barnesandnoble.com/w/stop-the-thyroid-madness-janie-a-bowthorpe/1101705814 [accessed 2015-03-05] [WebCite Cache ID 6WnJ1d16Z]

28. Lowe JC. Thyroid Science. 2009. Stability, Effectiveness, and Safety of Desiccated Thyroid vs Levothyroxine: A Rebuttal to the British Thyroid Association URL: http://www.thyroidscience.com/Criticism/lowe.3.16.09/bta.rebuttal.htm [accessed 2014-10-24] [WebCite Cache ID 6TYeR9lrC]

29. Lewis DH, Kumar J, Goulden P, Barnes DJ. Endocrine Abstracts. 2008. Improvements in quality of life in hypothyroid patients taking Armour thyroid URL: http://www.endocrine-abstracts.org/ea/0015/ea0015p359.htm [accessed 2015-08-17] [WebCite Cache ID 6argPAVOR]

30. PROAIR HFA Asthma Inhaler. URL: http://www.proairhfa.com/pro-air-hfa-inhaler/dose-counter.aspx [accessed 2015-03-05] [WebCite Cache ID 6WndjWUY5]

Abbreviations

ANCOVA: analysis of covariance
ANOVA: analysis of variance

FDA: Food and Drug Administration

Original Paper

Using Information and Communication Technologies for Family Communication and Its Association With Family Well-Being in Hong Kong: FAMILY Project

Man Ping Wang[1], PhD; Joanna TW Chu[2], PhD; Kasisomayajula Viswanath[3], PhD; Alice Wan[2], MSc; Tai Hing Lam[2], MD; Sophia S Chan[1], PhD

[1]The University of Hong Kong, School of Nursing, Hong Kong, China (Hong Kong)

[2]The University of Hong Kong, School of Public Health, Hong Kong, China (Hong Kong)

[3]Harvard TH Chan School of Public Health, Center for Community-Based Research, Dana-Farber Cancer Institute/Department of Social and Behavioral Sciences, Cambridge, MA, United States

Corresponding Author:
Tai Hing Lam, MD
The University of Hong Kong
School of Public Health
21 Sassoon Road, Pokfulam
Hong Kong,
China (Hong Kong)
Phone: 852 2819 9287
Fax: 852 2855 9528
Email: hrmrlth@hku.hk

Abstract

Background: Family communication is central to the family and its functioning. It is a mutual process in which family members create, share, and regulate meaning. Advancement and proliferation of information and communication technologies (ICTs) continues to change methods of family communication. However, little is known about the use of different methods for family communication and the influence on family well-being.

Objective: We investigated the sociodemographic factors associated with different methods of family communication and how they are associated with perceived family harmony, happiness, and health (3Hs) among Chinese adults in Hong Kong.

Methods: Data came from a territory-wide probability-based telephone survey using the Family and Health Information Trend survey (FHInTs). Frequency of family communication using different methods (ie, face-to-face, phone, instant messaging [IM], social media sites, and email) were recoded and classified as frequent (always/sometimes) and nonfrequent (seldom/never) use. Family well-being was measured using 3 questions of perceived family harmony, happiness, and health with higher scores indicating better family well-being. Adjusted odds ratios for family communication methods by sociodemographic characteristics and adjusted beta coefficients for family well-being by communication methods were calculated.

Results: A total of 1502 adults were surveyed. Face-to-face (94.85%, 1408/1484) was the most frequent means of communication followed by phone (78.08%, 796/1484), IM (53.64%, 796/1484), social media sites (17.60%, 261/1484), and email (13.39%, 198/1484). Younger age was associated with the use of phone, IM, and social media sites for family communication. Higher educational attainment was associated with more frequent use of all modes of communication, whereas higher family income was only significantly associated with more frequent use of IM and email (*P*=.001). Face-to-face (beta 0.65, 95% CI 0.33-0.97) and phone use (beta 0.20, 95% CI 0.02-0.38) for family communication were associated with significantly higher levels of perceived family well-being.

Conclusions: Socioeconomic disparities in using these information and communication technologies (ICT) methods for family communication were observed. Although traditional methods remain as the main platform for family communication and were associated with better family well-being, a notable proportion of respondents are using new ICT methods, which were not associated with perceived family well-being. Because ICTs will continue to diversify modes of family communication, more research is needed to understand the impact of ICTs on family communication and well-being.

XSL•FO
RenderX

Introduction

Family communication through both verbal and nonverbal interactions plays a central role in maintaining family relationships and enhancing family well-being [1]. It provides the foundation for family members (individuals who are related through biological, marital, cohabitation, and/or emotional bonding) to share meaning, to be connected, to be flexible in changing family rules, to achieve satisfaction, and to express and share attitudes, values, and beliefs [2,3]. Family communication includes the content (both verbal and nonverbal), frequency, and nature of family interactions, which defines the family and constructs family relationships [1]. Although a Western perspective often defines communication as an expression of "self" [4,5], a Chinese perspective defines communication as a way to develop and maintain personal relations and to reaffirm their membership in their respective social networks [6,7]. Nonconfrontational communication (harmony) is valued in Chinese culture and refers to expressing one's thoughts and feelings in an indirect and implicit manner [7]. Such pattern of communication is not only to preserve an individual's dignity, but also to protect family harmony and ties [7].

Family communication comprises an important part of several Western theories and models (eg, Family System Theory, Social Learning Theory, Olson's Circumplex Model, and McMaster Model of Family Functioning). Olson's Circumplex model [1] and the McMaster Model of Family Functioning [8] posit that a well-functioning family is characterized by positive communication which provides a basis for higher level of family cohesion and adaptability. Family communication has also been argued to be vital for family harmony, happiness, and health that underlie family well-being from a Chinese perspective [9,10].

In addition to the traditional means of communication, such as face-to-face and phone, new forms of information and communication technologies (ICTs), such as instant messaging (IM), social media sites (eg, Facebook, Twitter), and email, allow individuals to communicate and interact with one another [11,12]. Each medium has unique attributes and provides benefits otherwise not available from other means of communication. Face-to-face communication possesses nonverbal elements, instant feedback, complete identification, and real-time interaction [13]. The phone, although it lacks visual communication cues, provides instant feedback and real-time interaction across a wide geographical range. IM and email allow for words (and now photographs, videos, and audio clips) to be asynchronously exchanged among individuals and family groups. Social media sites allow for interconnectivity and provide an avenue for texts, photos, and video sharing [14]. ICTs also create a new pattern of family communication [12]; individuals are now able to perform multiple media tasks at the same time and interact with multiple individuals simultaneously. Compared with traditional forms of communication, ICTs are able to break barriers of time, space, location, and distance in making virtual communication accessible, feasible, and efficient [14,15]. Indeed, the rapid development of ICTs has changed and continues to transform the ways in which families interact and communicate [14,15].

Although a growing number of studies have examined communication behaviors (eg, pattern, frequency, and usage of ICTs) and interpersonal relationships and family functioning [12,14,15], it is uncertain whether ICTs enhance or weaken family relationships [14,15]. ICT use may increase the time families spend together, strengthen family bonds, improve family communication, and enable the maintenance of family relationships [16-20]. Other studies have suggested that quality family time has been significantly reduced and overuse of ICTs can lead to isolation from the family and failure to develop normal modes of expression, affecting the quality of family relationships [18,21-23].

Hong Kong, the most Westernized and urbanized city in China, is one of the most technologically advanced and connected cities in the world with ICTs readily integrated into the daily lives of Chinese people. Most households (78%) have personal computers at home connected to the Internet [24]. Fixed-line residential phone penetration exceeds 100% and mobile phone (mostly with multiple ICT functions) subscription plans penetrate more than 230% [24]. Therefore, Hong Kong provides an appropriate platform to understand how different modes of communication are used and the influence it has on family well-being. This study examines the use of ICTs for family communication and their influence on perceived family well-being.

Methods

Sampling

As part of the FAMILY Project, the Hong Kong Family and Health Information Trends Survey (FHInTS) was conducted from August 2012 to October 2012 using probability-based telephone surveys to collect information on general public opinions and behaviors on family health, information use, and health communication. Details of the survey design have been reported elsewhere [25,26]. In brief, a 2-stage random sampling method was used. First, telephone numbers were retrieved from residential telephone directories that covered approximately 76% of Hong Kong residents [24]. A computer program was used to generate a list of the telephone numbers in random order for interview. Invalid household numbers, nonresponses, and ineligible households (people aged <18 years or not able to speak Cantonese) were excluded. Second, after initial introduction of the study purpose by the interviewers, adult respondents were asked how many eligible individuals were living in the household. All eligible individuals were listed and the individual with the next birthday closest to the interview day was elected for interview. Each interview took approximately 25 minutes to complete. A total of 2127 individuals were eligible; of those, 1502 adults were successfully interviewed yielding a response rate of 70.62%. Ethics approval

was granted by the Institutional Review Board (IRB) of the University of Hong Kong/Hospital Authority Hong Kong West Cluster. Verbal informed consents were obtained and recoded verbatim, and the procedure was approved by the IRB.

Measurements

The definition of families (family members who are related through biological, marital, cohabitation, and/or emotional bonding) was explained to the respondents before asking questions about family communication and family well-being. The prevalence of different methods of communication was assessed by asking respondents how often each method (ie, face-to-face, phone, IM, social media sites, and email) was used to communicate with family with responses of "very often," "sometimes," "seldom," and "never." Perceived family harmony, happiness, and health (3Hs) are regarded as the main component of family well-being in Chinese society [9,10]. Family 3Hs were assessed by using 3 separate questions asking respondents to give a score from 0-10. Family well-being was calculated based on the composite score of the 3Hs with higher scores indicating better family well-being (possible total score ranged from 0-30). The test-retest reliability in another sample yielded an alpha=.81, showing that the scale was reliable over 1 month. The internal consistency of the scale was alpha=.84 for this sample. Socioeconomic status (SES) was measured using educational attainment, employment status, and monthly household income. Educational attainment was categorized as primary or below, secondary, and tertiary or above. Employment status was categorized as full-time, part-time, self-employed, and unemployed. Monthly household income was categorized as

≤HK $9999, HK $10,000-$19,999, HK $20,000-$29,999, HK $30,000-$39,999, HK $40,000-$59,999, and ≥HK $60,000 (US $1=HK $7.8).

Statistical Analysis

All data were weighted by sex and age using Hong Kong 2013 census data. Descriptive statistics were used to report prevalence of different methods of family communication. Associations of different methods to communicate with family by sex, age, marital status, and SES indicators (ie, income and educational attainment) were assessed by logistic regression, which yielded adjusted odds ratios (AORs) of family communication methods. Associations between perceived Family 3Hs (harmony, happiness, and health), family well-being, and different family communication methods were analyzed in a separate binary logistic regression model adjusting for sociodemographic characteristics. All analyses were conducted with SPSS 20 (SPSS Inc, Chicago, IL, USA).

Results

Of 1502 respondents, 54.51% (819/1502) were female, 73.39% (1103/ 1500) were aged between 25 and 64 years, and 63.33% (950/1500) were married or cohabitating (Table 1). Most respondents had secondary or greater education (85.60%, 1286/1502) and 63.24% (822/1300) had monthly family income of HK $20,000 or greater (average monthly income in Hong Kong was HK $20,200). Among 611 unemployed participants (40.66%, 611/1502), 45.9% (281/611) were retired and 35.1% (215/611) were homemakers.

Table 1. Sociodemographic characteristics of sample (N=1502).

Demographics	Unweighted, n (%)	Weighted, %
Sex		
Male	573 (38.15)	45.59
Female	929 (61.85)	54.51
Age [a]		
18-24	126 (8.40)	9.94
25-34	146 (9.73)	17.76
35-44	153 (10.20)	18.55
45-54	312 (20.80)	20.60
55-64	387 (25.80)	16.49
≥65	376 (25.07)	16.67
Marital status [b]		
Single	328 (21.88)	30.57
Married/cohabitating	1038 (69.25)	63.33
Other (divorced/widowed)	133 (8.87)	6.09
Educational attainment		
Primary or below	318 (21.17)	14.40
Secondary	733 (48.80)	47.68
Tertiary or above	451 (30.03)	37.92
Employment status		
Full-time	548 (36.48)	47.49
Part-time	135 (8.99)	8.77
Self-employed	38 (2.53)	3.09
Unemployed	781 (52.00)	40.66
Monthly household income (HK $) [c]		
≤9999	302 (23.65)	17.25
10,000-19,999	269 (21.06)	19.51
20,000-29,999	214 (16.76)	18.31
30,000-39,999	172 (13.47)	15.18
40,000-59,999	168 (13.16)	15.48
≥60,000	152 (11.90)	14.27

[a] Missing (unweighted=2, weighted=2).

[b] Missing (unweighted=3, weighted=2).

[c] Missing (unweighted=225, weighted=202).

The most frequent means of communication was through face-to-face (94.85%, 1408/1484) followed by telephone (78.08%, 1159/1484) and IM (53.64%, 796/1484) (Table 2). Some have also used social media sites (17.60%, 261/1484) and email (13.39%, 198/1484) for family communication. Although 63.18% (938/1484) of respondents reported never using email (followed by social media sites 59.65%, 885/1484), only a few (0.85%, 13/1484) reported never having communicated with family face-to-face.

Table 2. Prevalence (weighted) of different methods to communicate with family (N=1484).

Means of communication	Prevalence, n (%)			
	Very often	Sometimes	Seldom	Never
Face-to-face	1199 (80.77)	209 (14.08)	64 (4.30)	13 (0.85)
Mobile phone/phone	637 (42.91)	522 (35.18)	219 (14.74)	107 (7.18)
Instant messaging instruments (eg, WhatsApp, WeChat [WeiXin], LINE)	458 (30.84)	338 (22.80)	183 (12.31)	505 (34.05)
Social media sites (eg, Facebook, Twitter, Google+, WeiBo)	95 (6.43)	166 (11.17)	338 (22.75)	885 (59.65)
Email	53 (3.59)	145 (9.80)	348 (23.44)	938 (63.18)

Compared with males, females used IM more frequently (AOR 1.56, 95% CI 1.19-2.03) (Table 3). Younger age was associated with more frequent use of phone (P=.001), IM (P<.001), and social media sites (P<.001). Higher adjusted odds ratio of email use was observed for older age group (P=.03), particularly for respondents' aged 55 to 64 years of age (AOR 3.83, 95% CI 1.21-12.09). Compared with respondents with low education level (ie, primary or less), respondents with higher education had more frequent use of all different modes of communication. In particular, respondents with tertiary or greater level of education were strongly associated with IM (AOR 3.39, 95% CI 2.00-5.77) and email use (AOR 4.52, 95% CI 2.12-9.66). Higher household income was associated with more frequent IM (P<.001) and email use (P=.001). No association was observed for employment status with communication models.

Table 3. Associations of sociodemographic characteristics and use of different methods to communicate with family.[a]

Demographics	Face-to-face		Phone		Instant messaging		Social media sites		Email	
	AOR (95% CI)	*P*	AOR (95% CI)	*P*	AOR (95% CI)	*P*	AOR (95% CI)	*P*	AOR (95% CI)	*P*
Sex										
Male	1		1		1		1		1	
Female	0.68 (0.39, 1.20)	.19	1.29 (0.96, 1.73)	.10	1.56 (1.19, 2.03)	.001	1.11 (0.81, 1.51)	.52	1.28 (0.88, 1.87)	.19
Age (years)		.02[b]		.001[b]		<001[b]		<001[b]		.03[b]
18-24	1		1		1		1		1	
25-34	1.50 (0.45, 5.02)	.51	0.89 (0.46, 1.71)	.73	0.65 (0.39, 1.10)	.11	1.50 (0.84, 2.68)	.17	1.42 (0.49, 4.13)	.52
35-44	1.26 (0.34, 4.64)	.73	0.48 (0.24, 0.96)	.04	0.61 (0.34, 1.08)	.09	0.98 (0.51, 1.88)	.94	1.51 (0.49, 4.69)	.48
45-54	0.73 (0.20, 2.70)	.64	0.42 (0.21, 0.87)	.02	0.39 (0.21, 0.71)	.002	0.66 (0.33, 1.35)	.26	2.81 (0.91, 8.66)	.07
55-64	0.61 (0.16, 2.29)	.46	0.32 (0.15, 0.67)	.003	0.28 (0.15, 0.53)	<001	0.62 (0.29, 1.32)	.21	3.83 (1.21, 12.09)	.02
≥65	0.44 (0.11, 1.71)	.24	0.30 (0.14, 0.65)	.002	0.09 (0.04, 0.18)	<001	0.29 (0.12, 0.73)	.009	1.74 (0.49, 6.12)	.39
Marital status										
Single	1		1		1		1		1	
Married/co-habitating	2.83 (1.17, 6.86)	.02	1.68 (1.05, 2.68)	.03	1.94 (1.28, 2.93)	.002	1.03 (0.66, 1.61)	.90	3.52 (1.86, 6.65)	<.001
Others	0.93 (0.32, 2.68)	.90	2.28 (1.08, 4.83)	.03	1.38 (0.67, 2.82)	.38	0.63 (0.23, 1.70)	.36	2.24 (0.75, 6.67)	.15
Education		.005[b]		.03[b]		<001[b]		.43[b]		<001[b]
Primary or below	1		1		1		1		1	
Secondary	3.54 (1.80, 6.97)	<001	2.04 (1.35, 3.09)	.001	3.03 (1.90, 4.84)	<001	1.60 (0.84, 3.04)	.15	1.78 (0.89, 3.57)	.11
Tertiary or above	3.21 (1.32, 7.78)	.01	1.84 (1.12, 3.04)	.02	3.39 (2.00, 5.77)	<001	1.12 (0.55, 2.27)	.76	4.52 (2.12, 9.66)	<001
Employment										
Full-time	1		1		1		1		1	
Part-time	1.00 (0.36, 2.81)	.99	0.73 (0.43, 1.26)	.26	0.88 (0.55, 1.42)	.60	0.86 (0.47, 1.57)	.63	1.02 (0.52, 2.03)	.95
Self-employed	0.71 (0.13, 4.06)	.70	0.71 (0.34, 1.52)	.38	0.98 (0.48, 2.00)	.95	1.07 (0.48, 2.39)	.88	1.04 (0.45, 2.41)	.93
Unemployed	0.96 (0.46, 2.03)	.92	0.82 (0.56, 1.21)	.32	0.79 (0.57, 1.11)	.18	0.94 (0.63, 1.41)	.76	0.87 (0.55, 1.39)	.57
Family income (HK $)		.53[b]		.38[b]		<001[b]		.02[b]		.001[b]
≤9999	1		1		1		1		1	
10,000-19,999	1.24 (0.58, 2.67)	.58	1.24 (0.77, 2.01)	.38	1.28 (0.80, 2.03)	.31	0.86 (0.46, 1.61)	.64	1.83 (0.85, 3.95)	.12
20,000-29,999	2.13 (0.79, 5.75)	.14	0.67 (0.41, 1.10)	.11	1.48 (0.91, 2.40)	.11	0.95 (0.50, 1.80)	.89	1.76 (0.79, 3.91)	.16
30,000-39,999	2.61 (0.82, 8.32)	.10	0.94 (0.55, 1.62)	.82	2.29 (1.37, 3.82)	.002	1.14 (0.59, 2.19)	.70	1.81 (0.80, 4.13)	.16
40,000-59,999	0.85 (0.33, 2.14)	.73	1.31 (0.74, 2.32)	.36	2.84 (1.68, 4.81)	<001	1.64 (0.86, 3.13)	.13	2.71 (1.22, 5.99)	.01
≥60,000	2.68 (0.67, 10.60)	.16	1.43 (0.77, 2.68)	.26	3.63 (2.05, 6.40)	<001	1.39 (0.70, 2.76)	.35	2.94 (1.31, 6.59)	.009

[a]All variables were mutually adjusted. Frequency of use of communication methods was dichotomized as "very often/sometimes" and "seldom/never."
[b]*P* value

Frequent use of face-to-face communication was strongly associated with perceived family harmony (AOR 0.82, 95% CI 0.45-1.19), family happiness (AOR 0.59, 95% CI 0.22-0.96), family health (AOR 0.54, 95% CI 0.16-0.91), and overall family well-being (AOR 0.65, 95% CI 0.33-0.97) (Table 4). Similarly, more phone use was associated with family harmony (AOR

0.22, 95% CI 0.02-0.43), family happiness (AOR 0.36, 95% CI 0.15-0.56), and overall family well-being (AOR 0.20, 95% CI 0.02-0.38). Using new ICTs (ie, IM, social media sites, and email) was positively, but nonsignificantly, associated with family 3Hs and well-being.

Table 4. Association between family 3Hs and the use of different methods to communicate with family.[a]

Means of communication	Family harmony		Family happiness		Family health		Family well-being	
	Mean (SD)	Beta (95% CI)[b]	Mean (SD)	Beta (95% CI)[b]	Mean (SD)	Beta (95% CI)[b]	Mean (SD)	Beta (95% CI)[b]
Face-to-face								
Seldom/never	7.04 (2.31)	0	6.76 (2.32)	0	6.87 (2.17)	0	6.89 (2.01)	0
Always/sometimes	7.78 (1.48)	0.82 (0.45, 1.19)[e]	7.39 (1.52)	0.59 (0.22, 0.96)[d]	7.39 (1.50)	0.54 (0.16, 0.91)[d]	7.52 (1.29)	0.65 (0.33, 0.97)[e]
Phone								
Seldom/never	7.58 (1.77)	0	7.08 (1.92)	0	7.27 (1.72)	0	7.32 (1.55)	0
Always/sometimes	7.78 (1.46)	0.22 (0.02, 0.43)[c]	7.44 (1.45)	0.36 (0.15, 0.56)[d]	7.39 (1.49)	0.06 (–0.14, 0.27)	7.54 (1.27)	0.20 (0.02, 0.38)[c]
Email								
Seldom/never	7.70 (1.55)	0	7.32 (1.61)	0	7.37 (1.56)	0	7.47 (1.36)	0
Always/sometimes	7.98 (1.40)	0.14 (–0.10, 0.39)	7.62 (1.31)	0.11 (–0.15, 0.36)	7.35 (1.39)	–0.20 (–0.45, 0.05)	7.66 (1.18)	0.02 (–0.20, 0.23)
Instant messaging								
Seldom/never	7.70 (1.67)	0	7.28 (1.76)	0	7.29 (1.66)	0	7.43 (1.46)	0
Always/sometimes	7.77 (1.41)	0.01 (–0.17, 0.19)	7.43 (1.39)	0.06 (–0.13, 0.24)	7.43 (1.43)	–0.01 (–0.19, 0.17)	7.55 (1.22)	0.02 (–0.14, 0.17)
Social media sites								
Seldom/never	7.72 (1.57)	0	7.36 (1.62)	0	7.37 (1.57)	0	7.49 (1.38)	0
Always/sometimes	7.80 (1.35)	0.17 (–0.05, 0.38)	7.35 (1.32)	0.03 (–0.19, 0.24)	7.38 (1.36)	0.00 (–0.22, 0.21)	7.52 (1.11)	0.07 (–0.11, 0.26)

[a]Frequencies of use of communication methods were dichotomized as "very often/sometimes" and "seldom/never;" Family 3Hs ranged from 0 (totally unhealthy/unhappy/inharmonious) to 10 (very healthy/happy/harmonious), with 5 indicating "half-half."

[b] Adjusting for sex, age, education attainment, monthly household income, and marital status.

[c]$P<.05$

[d]$P<.01$

[e]$P<.001$

Discussion

Although research on the interplay between technological advancements and family functioning are needed [11,15,27] and increasingly reported, little consensus has been found on the impact of ICTs on family well-being. Findings on Chinese population are scarce and our study provides the first evidence on ICTs use and perceived family well-being among Chinese adults.

Overall, the findings are consistent with studies elsewhere [12,28,29] that showed traditional means of communication (ie, face-to-face and phone) were most frequently endorsed compared with emerging ICTs (ie, IM, social media, and email). Our findings on the associations between sociodemographic characteristics and use of different communication methods also revealed similar trends and characteristics to those in the literature [30,31]. Younger individuals would have grown up in a generation riddled with new technology and are, therefore, more likely to embrace ICTs in various forms. Higher education level and higher household income were associated with more frequent ICT use (ie, IM and email). One possible explanation is that individuals with higher education were more likely to be professional workers or employed in office settings that have greater access to computers and mobile phones that allowed for IM and email use. Similarly, those with higher household income may have greater resources (eg, financially) and accessibility (eg, Internet connection at home, work, and mobile subscription plans) to ICTs.

Of significance is that traditional methods of communication (face-to-face and phone) were strongly associated with higher levels of perceived family well-being. This finding is especially important given that recent studies suggested a transformative trend toward more frequent use of ICTs than traditional methods of communication, particularly in the younger age group [15,18,30,32]. Specifically, we found that face-to-face communication was significantly related to all 3 dimensions of

family well-being (harmony, happiness, and health). Using the phone as a communication method was also associated with higher levels of family harmony, happiness, and family well-being. One explanation is that the quality of communication through face-to-face and phones are richer than those of ICTs and, thus, provide greater communication satisfaction [28]. Face-to-face communication conveys verbal, nonverbal, and social context cues simultaneously and provides immediate and synchronized feedback. These are all fundamental qualities to establishing human relationships that, in turn, affect family well-being [33]. Particularly in Chinese societies, where the style of communication is often indirect (messages are often implicit and the meanings are to be inferred from contextual cues), face-to-face communication represents an important means of communication. Although not directly examining family well-being, another study reported similar findings on ICT use and perceived adolescent well-being [34]. More frequent use of social media sites by college students to communicate with their parents was more strongly associated with self-reported loneliness. On the other hand, frequent phone communication was associated with more positive qualities in the parent-child relationship, including greater satisfaction, intimacy, support, and instrumental aid [34].

Given the importance of face-to-face communication, our findings support the notion that ICTs should not replace traditional methods of communication, but rather should be utilized as a supplement. Studies also found when ICTs were used as a substitute, the effects on interpersonal relationships were negative [35]. The absence of nonverbal cues and tacit knowledge makes communication difficult and hinders relationship formation, cohesion, and trust [36]. For example, the lack of social presence creates an environment in which individuals easily misinterpret emotions and/or make incorrect assumptions. However, when used as a complement to face-to-face communication, ICTs facilitated the maintenance of interpersonal relationships [35]. These studies along with our findings suggest the importance of ICTs supplementing traditional methods of communication on improving family well-being. Therefore, efforts to improve family communication and well-being need to focus on informing individuals about the importance of face-to-face communication and the opportunities and pitfalls that ICTs bring.

One of the limitations of this study is the broad categories used in assessing the different mediums of communication. For example, we did not differentiate between fixed-line residential phone and mobile phone. We also did not incorporate videoconference services, such as Skype and Jaber, which at the time were seldom used. Another limitation is that we did not assess the geographical distance between family members, which can have an influence on the choice of communication methods [17]. None-the-less, others found that compared with other modes of communication, face-to-face communication showed a strong positive relationship to frequency of contact after controlling for locality [12]. The cross-sectional design cannot be used to determine causality. The sampling method only covered residential telephone directories; therefore, households that used mobile phones only were excluded. Finally, data from individuals younger than 18 years were not collected. Given that children and adolescents are more deeply immersed in the digital world, examining their behaviors and pattern of use may provide a clearer picture to how ICTs impact family well-being across different life spans.

This study suggests several avenues for future research. The quality of communication has rarely been measured [15,37]. The context and content of the dialog may provide more insight as to the quality of the communication that is likely to influence family relationships and family well-being more strongly. Future studies are also needed to examine the diverse range of ICTs, from preference and pattern of use to its association and causality.

Although traditional methods remained as the main platform for communication within the family and were associated with better family well-being, a notable proportion were using new ICT methods (ie, IM, social media sites, and email). Socioeconomic disparities in using these ICT methods for family communication were observed. Because ICTs will continue to diversify modes of family communication, more research is needed to understand the impact of ICTs on family communication and well-being.

Conflicts of Interest

None declared.

References

1. Olson D. Circumplex model of marital and family systems. J Fam Ther 2000;22:144-167.
2. Pearson JC. Communication in the Family: Seeking Satisfaction in Changing Times. New York: HarperCollins College Publishers; 1993.
3. Galvin KM. In: Bylund CL, Brommel B, editors. Family Communication: Cohesion and Change. Boston: Allyn and Bacon; 2004.
4. Katriel T, Philipsen G. "What we need is communication": "Communication" as a cultural category in some American speech. Communication Monographs 2009 Jun 02;48(4):301-317. [doi: 10.1080/03637758109376064]

XSL•FO

RenderX

5.   Carbaugh D, van Over B. Interpersonal pragmatics and cultural discourse. J Pragmat 2013;58:142-145.

6.   Yum J. The impact of Confucianism on interpersonal relationships and communication patterns in east Asia. Communication Monographs 1988 Dec;55(4):374-388. [doi: 10.1080/03637758809376178]

7.   Chang H. Harmony as performance: the turbulence under Chinese interpersonal communication. Discourse Studies 2001 May 01;3(2):155-179. [doi: 10.1177/1461445601003002001]

8.   Epstein N, Bishop D, Levin S. The McMaster Model of Family Functioning. J Marital Fam Ther 1978;4:19-31. [doi: 10.1111/j.1752-0606.1978.tb00537.x]

9.   Lam WW, Fielding R, McDowell I, Johnston J, Chan S, Leung GM, et al. Perspectives on family health, happiness and harmony (3H) among Hong Kong Chinese people: a qualitative study. Health Educ Res 2012 Oct;27(5):767-779 [FREE Full text] [doi: 10.1093/her/cys087] [Medline: 22907531]

10. Chan SS, Viswanath K, Au DW, Ma CM, Lam WW, Fielding R, et al. Hong Kong Chinese community leaders' perspectives on family health, happiness and harmony: a qualitative study. Health Educ Res 2011 Aug;26(4):664-674 [FREE Full text] [doi: 10.1093/her/cyr026] [Medline: 21536713]

11. Blinn-Pike L. Technology and the family: an overview from the 1980s to the present. Marriage & Family Review 2009 Oct 27;45(6-8):567-575. [doi: 10.1080/01494920903224459]

12. Stern M, Messer C. How family members stay in touch: a quantitative investigation of core family networks. Marriage & Family Review 2009 Oct 27;45(6-8):654-676. [doi: 10.1080/01494920903224236]

13. Chesebro J. Computer-Mediated Communication: Human Relationships in a Computerized World. Tuscaloosa, AB: University of Alabama Press; 1989.

14. Rudi J, Dworkin J, Walker S, Doty J. Parents' use of information and communications technologies for family communication: differences by age of children. Information, Communication & Society 2014 Jul 02;18(1):78-93. [doi: 10.1080/1369118X.2014.934390]

15. Carvalho J, Francisco R, Relvas A. Family functioning and information and communication technologies: How do they relate? A literature review. Computers in Human Behavior 2015 Apr;45:99-108. [doi: 10.1016/j.chb.2014.11.037]

16. Lanigan J. A sociotechnological model for family research and intervention: how information and communication technologies affect family life. Marriage & Family Review 2009 Oct 27;45(6-8):587-609. [doi: 10.1080/01494920903224194]

17. Bacigalupe G, Lambe S. Virtualizing intimacy: information communication technologies and transnational families in therapy. Fam Process 2011 Mar;50(1):12-26. [doi: 10.1111/j.1545-5300.2010.01343.x] [Medline: 21361921]

18. Williams AL, Merten MJ. iFamily: Internet and social media technology in the family context. Fam Consum Sci Res J 2011;40:150-170. [doi: 10.1111/j.1552-3934.2011.02101.x]

19. Chesley N, Fox B. E-mail's use and perceived effect on family relationship quality: variations by gender and race/ethnicity. Sociological Focus 2012 Jan;45(1):63-84. [doi: 10.1080/00380237.2012.630906]

20. Kanter M, Afifi T, Robbins S. The impact of parents "friending" their young adult child on facebook on perceptions of parental privacy invasions and parent-child relationship quality. J Commun 2012 Sep 14;62(5):900-917. [doi: 10.1111/j.1460-2466.2012.01669.x]

21. Nie NH. Sociability, interpersonal relations, and the Internet: reconciling conflicting findings. American Behavioral Scientist 2001 Nov 01;45(3):420-435. [doi: 10.1177/00027640121957277]

22. Bargh J. Beyond simple truths: the human-Internet interaction. J Social Isssues 2002 Jan;58(1):1-8. [doi: 10.1111/1540-4560.00245]

23. Turkle S. Alone Together: Why We Expect More from Technology and Less from Each Other. New York: Basic Books; 2011.

24. Full Report on CSL Service Interruption on 13 May 2013. URL: http://www.ofca.gov.hk/filemanager/ofca/en/content_723/csl_report_20130516_full.pdf [accessed 2015-08-18] [WebCite Cache ID 6arYpoyjI]

25. Wang MP, Viswanath K, Lam TH, Wang X, Chan SS. Social determinants of health information seeking among Chinese adults in Hong Kong. PLoS One 2013;8(8):e73049 [FREE Full text] [doi: 10.1371/journal.pone.0073049] [Medline: 24009729]

26. Wang MP, Wang X, Viswanath K, Wan A, Lam TH, Chan SS. Digital inequalities of family life information seeking and family well-being among Chinese adults in Hong Kong: a population survey. J Med Internet Res 2014;16(10):e227 [FREE Full text] [doi: 10.2196/jmir.3386] [Medline: 25281831]

27. Aponte R. The communications revolution and its impact on the family: significant, growing, but skewed and limited in scope. Marriage & Family Review 2009 Oct 27;45(6-8):576-586. [doi: 10.1080/01494920903396778]

28. Baym N, Zhang Y, Lin M. Social interactions across media: interpersonal communication on the internet, telephone and face-to-face. New Media & Society 2004 Jun 01;6(3):299-318. [doi: 10.1177/1461444804041438]

29. Kim H, Kim G, Park H, Rice R. Configurations of relationships in different media: FtF, email, instant messenger, mobile phone, and SMS. J Comput Mediat Commun 2007;12:1183-1207. [doi: 10.1111/j.1083-6101.2007.00369.x]

30. Siibak A, Vittadini N. Editorial: introducing four empirical examples of the "generationing" process. CP 2012;6(2):1. [doi: 10.5817/CP2012-2-1]

31. James BD, Boyle PA, Yu L, Bennett D. Internet use and decision making in community-based older adults. Front Psychol 2013;4:1-10. [doi: 10.3389/fpsyg.2013.00605]

XSL•FO

RenderX

32.   Zhong B. From smartphones to iPad: Power users' disposition toward mobile media devices. Computers in Human Behavior 2013 Jul;29(4):1742-1748. [doi: 10.1016/j.chb.2013.02.016]

33.   Ramirez A, Broneck K. `IM me': instant messaging as relational maintenance and everyday communication. Journal of Social and Personal Relationships 2009 Jul 31;26(2-3):291-314. [doi: 10.1177/0265407509106719]

34.   Ramsey MA, Gentzler AL, Morey JN, Oberhauser AM, Westerman D. College students' use of communication technology with parents: comparisons between two cohorts in 2009 and 2011. Cyberpsychol Behav Soc Netw 2013 Oct;16(10):747-752. [doi: 10.1089/cyber.2012.0534] [Medline: 23679572]

35.   Kraut R, Kiesler S, Boneva B, Cummings J, Helgeson V, Crawford A. Internet paradox revisited. J Social Issues 2002 Jan;58(1):49-74. [doi: 10.1111/1540-4560.00248]

36.   Powell A, Piccoli G, Ives B. Virtual teams. SIGMIS Database 2004 Feb 03;35(1):6-36. [doi: 10.1145/968464.968467]

37.   Valkenburg PM, Peter J, Schouten AP. Friend networking sites and their relationship to adolescents' well-being and social self-esteem. Cyberpsychol Behav 2006 Oct;9(5):584-590. [doi: 10.1089/cpb.2006.9.584] [Medline: 17034326]

Abbreviations

3H: perceived family harmony, happiness, and health
AOR: adjusted odds ratio
FHInT: Family and Health Information Trend
ICT: information and communication technologies
IM: instant messaging

Original Paper

The Effects of Web-Based Patient Access to Laboratory Results in British Columbia: A Patient Survey on Comprehension and Anxiety

Geneviève Mák[1*], MSc; Heather Smith Fowler[1*], MA; Chad Leaver[2*], MSc; Simon Hagens[2*]; Jennifer Zelmer[2], PhD

[1]Social Research and Demonstration Corporation, Ottawa, ON, Canada

[2]Canada Health Infoway, Toronto, ON, Canada

[*]these authors contributed equally

Corresponding Author:
Simon Hagens
Canada Health Infoway
150 King St West, suite 1300
Toronto, ON, M5H1J9
Canada
Phone: 1 4164274182
Fax: 1 4165935911
Email: shagens@infoway-inforoute.ca

Abstract

Background: Web-based patient access to personal health information is limited but increasing in Canada and internationally.

Objective: This exploratory study aimed to increase understanding of how Web-based access to laboratory test results in British Columbia (Canada), which has been broadly available since 2010, affects patients' experiences.

Methods: In November 2013, we surveyed adults in British Columbia who had had a laboratory test in the previous 12 months. Using a retrospective cohort design, we compared reported wait-time for results, test result comprehension, and anxiety levels of "service users" who had Web-based access to their test results (n=2047) with those of a general population panel that did not have Web-based access (n=1245).

Results: The vast majority of service users (83.99%, 95% CI 82.31-85.67) said they received their results within "a few days", compared to just over a third of the comparison group (37.84%, 95% CI 34.96-40.73). Most in both groups said they understood their test results, but the rate was lower for service users than the comparison group (75.55%, 95% CI 73.58-77.49 vs 84.69%, 95% CI 82.59-86.81). There was no significant difference between groups in levels of reported anxiety after receiving test results.

Conclusions: While most patients who received their laboratory test results online reported little anxiety after receiving their results and were satisfied with the service, there may be opportunities to improve comprehension of results.

Introduction

The use of consumer health solutions, including Web-based patient access to laboratory test results, is limited but expanding internationally [1]. In Canada, more than 8 in 10 adults express interest in use of such services, but in 2014, only about 6% said that they had online access to their laboratory test results. In British Columbia, availability was much higher than the national average (27% of those surveyed). Web-based access to laboratory test results has been available in most regions there since 2010 [2]. Subscribers to the direct lab access service in British Columbia create an account and register with secure passwords; access to the service is free. Patients often learn about the service and are provided secure access through the lab facility. After a lab test has been conducted, service users obtain Web-based access to their test results. The lab report is presented verbatim, as the ordering clinician would receive it, without additional information. Figure 1 presents a screenshot of a typical lab report, as would be available to patients.

Ordering clinicians may contact their patients about the results or not, based on their clinical practice. Physicians do not have to subscribe to the service for patients to have access to lab test results.

As a relatively new technology, there is sparse literature about the benefits and risks of direct patient access to lab test results online, especially results that are abnormal or require follow-up with a health care provider [3]. For example, there is a lack of consensus on (1) best practices in direct patient notification of abnormal results, (2) whether patients will know what to do with the results, and (3) how they will react if they receive abnormal results online. Concerns about potential risks, such as patient anxiety or confusion, have been documented in the literature [4]. However, this association has not been found in the small body of research to date in this area. On the other hand, previous studies have identified advantages for patients who accessed their personal health records, such as improved quality of interactions with physicians, motivation to be better informed about and manage their own health [5-7], as well as a reduction of outpatient visits [8,9].

To our knowledge, this quasi-experimental study is the first to assess the effects of direct patient access to medical laboratory tests in Canada. Results presented here pertain to the experiences of British Columbia patients after seeing their lab test results online, focusing specifically on comprehension of test results and reported anxiety.

Figure 1. Screen capture of a sample lab report.
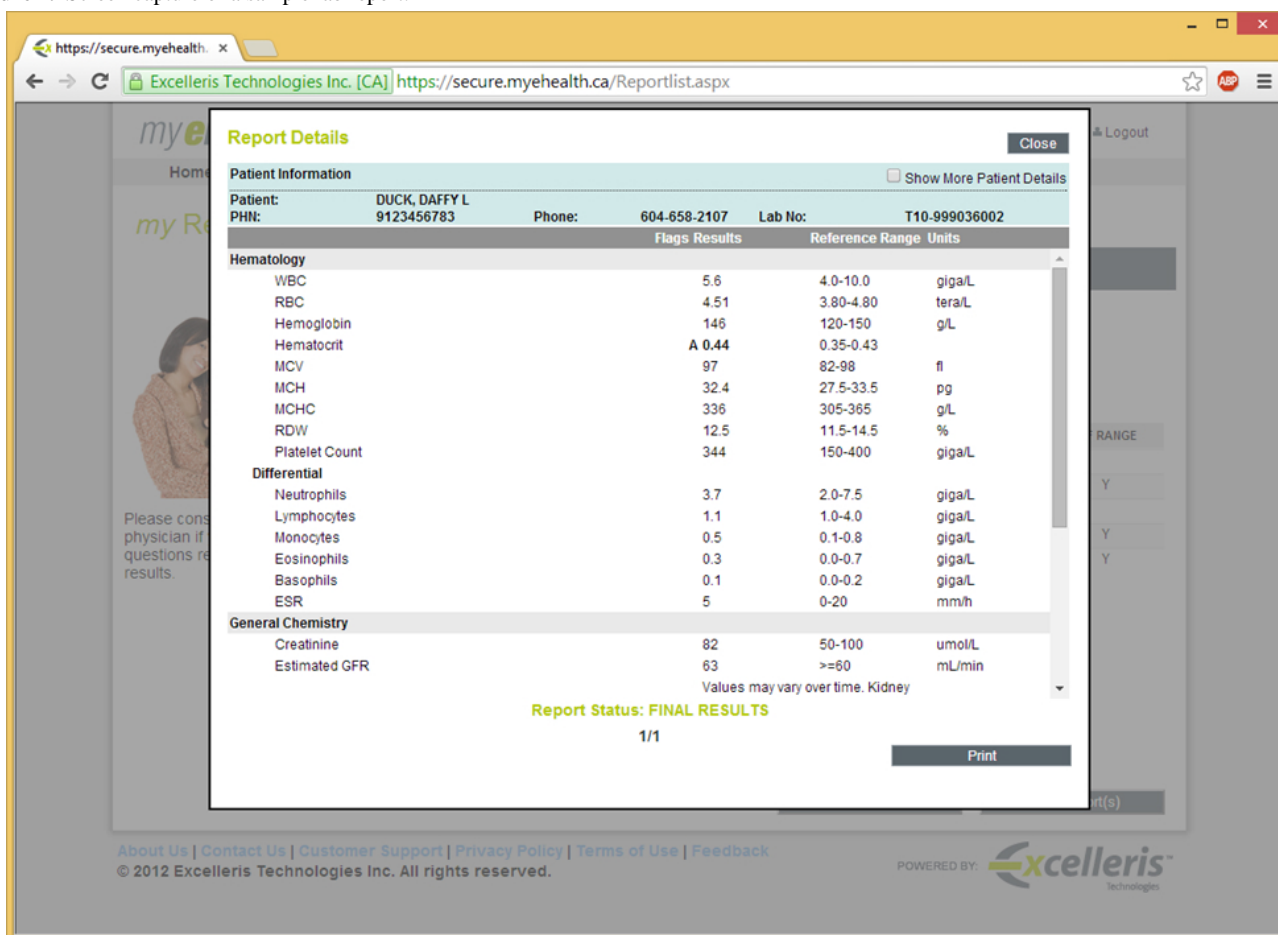
Methods

Data Sources

The service user and comparison cohorts were recruited separately. Service users were recruited from the subscriber database of the British Columbia service provider, specifically, from the approximately 15,000 subscribers who had given prior consent to be contacted for research purposes. Invitation emails were sent directly by the service provider in November 2013 to randomly selected subscribers (n=11,300) who were 18 years of age or older and had had a medical lab test conducted in the previous 12 months, the results of which they had accessed online. A pre-test of the survey was conducted among a small sample of participants in the online service users cohort (n=24). A total of 2047 service users fully completed the survey, for a response rate of 18.12%. While the bulk of the service users group consisted of participants who had first received their most recent test results online, some subscribers had first learned their most recent result in-person from their health care professional. The latter were omitted from some of the analyses to isolate the impact of receiving results online.

The comparison cohort was recruited in December 2013 from members of a general population panel maintained by Vision Critical, the composition of which is benchmarked against known census subgroups by age, gender, region, education, and income. The Vision Critical panel consists of approximately

130,000 panelists, of whom 15,000 reside in British Columbia. Each month, the panel fields a monthly survey to keep information about panelists current, to pre-screen panelists for specific study objectives, and to keep them actively engaged in the panel. In this case, the monthly screening tool served both to identify a potential comparison group and to target recruitment to match the online service users group as closely as possible on characteristics of age and gender. Two questions were used to pre-screen participants for our study: (1) "Have you had any medical laboratory tests conducted in the past 12 months?", and (2) "If yes, how did you receive the results for your medical laboratory test(s): in-person, online, via mail, email, or over the phone?" The 20.60% of those screened who reported having received results online in the past 12 months were disqualified from our study. A total of 2762 panelists were recruited for the comparison group; these were randomly selected from demographic subgroups to balance the study cohorts. Of these, 1245 people fully completed the survey, resulting in a response rate of 45.08%.

Because of the relatively low response rate of service users, the results obtained are not considered representative of the broader population of online service users. The differences in response rates between the service users group and the comparison group may also have resulted in response bias (ie, nonresponse and voluntary response biases). Although statistical methods such as analytical weighting were used to balance observable characteristics of the two samples, differences in measured outcome indicators can be confounded with unobservable factors, such as familiarity with lab results or anxiety associated with inexperience. The results of this study should be therefore be interpreted with some caution. This study received ethics approval from Institutional Review Board Services (IRBS), Canada.

Measures

Our survey included four sections: (1) self-reported health status and laboratory testing needs, (2) experience receiving the most recent lab test result, (3) perception of Web-based access to laboratory results (service users group only), and (4) sociodemographic characteristics and access to online consumer health solutions. Two questions were asked regarding respondents' comprehension of lab test results: (1) "Was it clear if you needed to follow-up with your doctor? (Yes/No)", and (2) "How confident are you that you fully understood your lab results?" on a scale of 1-10 (where 1=Not at all confident/10=Extremely confident).

The Global Anxiety-Visual Analog Scale (GA-VAS) (where "0" means "not at all anxious" and "100" means "extremely anxious") was used because it has been reported to be useful and valid in assessing anxiety as a single construct (with many perceptible gradations) and as a measure of anxiety at a specific point in time, such as pre-operative anxiety [10].

Analysis

Analyses were conducted using SPSS version 18.0. Preliminary analysis of our sample showed that the service users group had significantly more abnormal test results than the comparison group (633/1806, 35.1% vs 144/897, 16.1%). Since this variable was seen as key to patients' overall experience, we weighted the sample on this variable in order to adjust for its effect (see Multimedia Appendix 1). Independent Student's t tests for means or proportions of responses were used to assess differences between the two cohorts on sociodemographics, comprehension, and anxiety. For categorical responses, independent Student's t tests were used for inference instead of distributional tests such as chi-square tests, in order to be able to detect any substantial differences across the specific response items. In addition, logistic regression techniques were used to examine the association between sociodemographic variables and comprehension.

It should be noted that missing responses were excluded from analyses. Also, in each table, n is calculated based on the actual number of respondents in the sample, while proportions are estimated with analytical weight.

Results

Sample Profile

Although the two cohorts were balanced in terms of age and gender, there were some statistically significant differences in other sociodemographic characteristics and health status (see Table 1). For example, the service users group included fewer participants who were born in Canada and who spoke English at home, more participants with university education, more urban participants, and more who made over CAN $100,000 in annual income. In terms of health status, participants from the service users group reported being in slightly poorer health than the comparison group and having undergone more medical lab tests (3 or more times) in the last 12 months.

Our overall sample included a high proportion of participants over age 55 who had a chronic illness and who were therefore more likely to require health care services. Women also outnumbered men, consistent with existing literature, which has found that more women subscribe to health portals than men [11,12].

Outcome analyses were conducted...at the time of the survey(1862/2047...respectively), using the weighting procedures described in Multimedia Appendix 1.

XSL•FO

RenderX

Table 1. Characteristics of survey participants.

	Service users group (n=2047)		Comparison group (n=1245)		P
	n	%	n	%	
Gender					
Male	770	38.08	474	38.21	
Female	1252	61.92	767	61.79	
Age group, years					
18-34	193	9.75	115	9.41	
35-54	566	28.60	347	28.27	
55+	1220	61.65	765	62.32	
Immigration status and language spoken at home					
Born in Canada: yes	1429	70.29	1043	83.84	<.001
Language speak at home: English	1938	95.14	1231	98.91	<.001
Income, CAN$					
<$50K	545	33.64	429	42.66	<.001
$50K-$100k	626	38.64	369	36.74	
≥$100K	449	27.72	207	20.60	<.001
Education					
High school or less	164	8.18	155	12.76	<.001
Some/Completed college	708	35.29	515	42.30	<.001
Some university +	1134	56.53	547	44.93	<.001
Region					
Greater Vancouver	1107	54.45	355	28.56	<.001
Vancouver Island	442	21.74	287	23.12	
BC Southern Interior	340	16.72	446	35.88	<.001
BC Northern Interior	23	1.11	66	5.28	<.001
Overall health (last 12 months)					
Excellent/Very good	868	42.48	552	44.42	
Good	710	34.75	445	35.78	
Fair/Poor	465	22.76	246	19.79	<.05
Has a chronic health condition					
Yes	1213	60.35	727	58.82	
Number of medical lab tests conducted (past 12 months)					
≥6 times	450	20.84	148	11.75	<.001
3-5 times	834	40.55	412	33.05	<.001
2 times	440	21.91	363	29.30	<.001
1 time	323	16.70	322	25.90	<.001

Wait Time to Receive Results

As expected, the wait time to receive lab test results was considerably shorter for the service users group, 87.50% (1624/1856) of whom first learned the result of their most recent lab test online. The majority of service users (83.99%, 95% CI 82.31-85.67) said they waited only "a few days" following their lab test before receiving their results, compared to just over a third of the comparison group (37.84%, 95% CI 34.96-40.73). Table 2 presents more information about the wait time to receive lab test results.

XSL•FO
RenderX

Table 2. Wait time to receive lab results.

	Service users group (n=1818), n (%) 95% CI	Comparison group (n=1087), n (%) 95% CI
Received results within a few days	1527 (83.99[a]) 82.31-85.67	411 (37.84) 34.96-40.73
Received results in about a week	209 (11.50) 10.03-12.96	335 (30.85 [a]) 28.11-33.61
Received results in 1-2 weeks	56 (3.08) 2.29-3.89	221 (20.32 [a]) 17.92-22.71
Received results between 2 and 4 weeks	16 (0.88) 0.45-1.31	85 (7.81 [a]) 6.21-9.40
Received results in 4 weeks	10 (0.55) 0.21-0.89	35 (3.18 [a]) 2.14-4.23

[a]$P<.001$

Comprehension of Lab Test Results

All those who knew their most recent test results (n=2990) were asked about their confidence in fully understanding the results, as measured by a score of 7 or higher on a scale of 1-10. The majority of both service users and the comparison group were confident they understood the test results themselves, but the percentage was lower for service users (see Table 3).

Table 3. Comprehension of lab results.

All who received results	Service users group (n=1852), n (%) 95% CI	Comparison group (n=1119), n (%) 95% CI
How confident are you that you fully understood your lab results (score=7-10)?	1399 (75.55) 73.58-77.49	948 (84.69[a]) 82.59-86.81

[a]$P<.001$

To further explore what might influence patients' comprehension of their lab results, we conducted a logistic regression using sociodemographic and health service–related variables, and adjusted the results for receipt of abnormal lab test results. The variable "first learned the result online" was used this time to assess how comprehension was influenced by the service itself and not simply by being a subscriber to the service more generally. As expected, first learning test results online was a significant *negative* predictor of comprehension, as were younger ages and lower levels of education (see Table 4).

Table 4. Very confident in fully understanding lab results (logistic regression; n=2796; 194 cases were excluded from analysis; % correct predicted values: 79.2%).

	B (SE)	95% CI for exp b		
		Lower	exp b	Upper
Gender				
Female	-0.16 (0.10)	0.70	0.85	1.04
Male (ref)	-	-	-	-
Education				
High school or under	-0.72 (0.15)[a]	0.36	0.49	0.66
Some/Completed college	0.24 (0.10)[b]	0.65	0.79	0.97
Some university and + (ref)	-	-	-	-
Age				
18-34	-0.48 (0.16)[a]	0.45	0.49	0.85
35-54	-0.27 (0.11)[a]	0.62	0.77	0.95
55 and + (ref)	-			
First learned the result online				
Yes	-0.57 (0.10)[a]	0.47	0.57	0.69
No (ref)	-	-	-	-
Having a chronic condition				
Yes	-0.20 (0.10)[c]	0.68	0.82	1.00
No (ref)	-	-	-	-
Number of lab tests conducted (past 12 months)				
≥6 times	0.20 (0.13)	0.95	1.22	1.57
≤5 (ref)	-	-	-	-
Constant	2.15 (0.13)[a]		8.58	

[a]$P<.001$

[b]$P<.05$

[c]$P<.01$

Anxiety

We also conducted between-group analyses on respondents' reported level of anxiety *after* receiving their lab test results. Since the distribution of GA-VAS scores was positively skewed, we divided these into three categories: no reported anxiety (0), "low" anxiety (1-49), and "some" anxiety (50-100). To isolate the influence of Web-based access, we omitted from the analysis service users who had received the results of their most recent test in-person from their doctor or usual place of care.

We found no significant differences between service users and the comparison group in their level of anxiety following receipt of test results; as always, we adjusted for the effect of having a test result out of the normal range. As seen in Table 5, the majority of patients in both cohorts reported low or no anxiety after receiving test results. Table 5 also shows the results of analysis with a subgroup of participants in both cohorts who had a chronic health condition. Here, differences between the two cohorts emerged, such that service users reported being significantly less likely to be anxious at both ends of the scale (none and some).

Table 5. Anxiety after receiving lab test results.

	All		Subgroup with chronic condition	
	Subsample of service users who first learned results online (n=1478), n (%) 95% CI	Subsample comparison group who learned otherwise (n=1312), n (%) 95% CI	Subsample of service users who first learned results on-line (n=881), n (%) 95% CI	Subsample comparison group who learned otherwise (n=779), n (%) 95% CI
No anxiety (0)	540 (36.54) 34.08-38.99	447 (34.05) 31.49-36.62	310 (35.19[a]) 32.03-38.34	239 (30.64) 27.40-33.89
Low anxiety (1-49)	734 (49.66) 47.11-52.22	670 (51.07) 48.37-53.78	453 (51.42) 48.12-54.72	401 (51.47) 47.96-54.98
Some anxiety (50-100)	204 (13.80) 12.04-15.56	195 (14.87) 12.95-16.80	118 (13.39) 11.14-15.64	139 (17.89[a]) 15.19-20.58

[a]$P<.05$

We also explored the link between anxiety and comprehension among service users specifically, to determine if individuals who first learned their results online *and* who reported lower levels of comprehension also reported more anxiety. To measure comprehension in relation to anxiety, respondents were asked how clear it was they needed to follow-up with their health care provider. As expected, service users who first learned the results of their most recent lab test online and who indicated it was clear they needed to follow-up were *less* likely to report no anxiety (38.30%, 95% CI 35.44-41.16) as those who reported not being clear about the need for follow-up (29.84%, 95% CI 24.69-34.98). The same pattern held at the other end of the anxiety spectrum. Overall, Table 6 shows that participants who were not clear on the need for follow-up were more likely to report being anxious.

Table 6. Anxiety level after receiving lab test results according to comprehension.

	Service users who first learned results online (n=1412)	
	Clear if you need to follow-up? Yes (n=1107) n (%) 95% CI	Clear if you need to follow-up? No (n=305) n (%) 95% CI
No anxiety (0)	424 (38.30[a]) 35.44-41.16	91 (29.84) 24.69-34.98
Low (1-49)	551 (49.77) 46.83-52.72	153 (50.16) 44.54-55.78
Some (50-100)	132 (11.92) 10.01-13.83	61 (20.00[a]) 15.50-24.50

[a]$P<.001$

Discussion

Principal Findings

Our findings suggest that patient experience overall can be improved by the availability of Web-based lab results, but with important caveats. Service users were more likely to report a shorter wait for test results and high levels of satisfaction with the online service. There was no overall difference in post-result anxiety levels between those who saw results online and those who received results in other ways (eg, by mail or telephone), although among the subset of patients with chronic conditions, service users were less likely to report anxiety. However, service users were *more* likely to report lower comprehension of lab test results than the comparison group, and there was a significant correlation between anxiety and lower comprehension. This is not surprising, given that the format of lab results provided by the Web-based platform is the same as that provided to ordering clinicians, with no additional contextual or explanatory information for patients.

Our study results support Pyper et al's call for more information and tools to help patients understand and interpret their health information [5]. A range of tools has been suggested in the literature, including a glossary, integration with other health records, and patient education/information support.

Limitations

Although our comparison sample was recruited from a general population panel, differences between the service users group and the comparison group—including the rate of abnormal test results—somewhat limits the study's external validity. The service user group also had a lower response rate at 18%, which may indicate a possible response bias. Both cohorts were also recruited online, so our findings may not apply to the rapidly diminishing proportion of the population that does not have Internet access and a degree of digital literacy.

We also do not know much about the sequence and timing involved in obtaining test results by different means, and follow-up information and support. For example, we could not differentiate between tests conducted for diagnostic purposes versus for monitoring a previously known health condition. Additional clinical information would have provided a more focused interpretation of results about the patient experience. While we deliberately focused many of our survey questions on the most recent test to enhance precision of responses (ie, content validity) and their reliability, patients' most recent tests may not reflect their typical experience with lab tests or with

direct lab access in general. Moreover, we did not focus our study on the experiences of patients with abnormal test results, possibly diluting any negative effects of Web-based access on anxiety and comprehension, if they exist.

Finally, while we balanced our two cohorts as closely as possible and weighted the comparison cohort to take into account the type of test result received (ie, in the normal range or not), the quasi-experimental design of our study limits our ability to attribute observed differences to the intervention with the same degree of confidence as with random assignment to treatment, had that been possible.

Conclusions

Laboratory tests are among the most common interventions in modern health systems, and effective communication of test results and required follow-up is a priority for research. As jurisdictions around the world move toward widespread adoption of digital health technologies for patients, better understanding is needed of the effects of such services on both patients and health care practice. This study explored these issues in relation to one such technology—direct patient access to Web-based lab test results—currently in use in British Columbia, Canada. While exploratory, our study supports the emerging literature suggesting that personal health records are positively received by patients [13,14]. It also supports a recent systematic review that found access to health records reduced or had no effect on anxiety [15]. In our case, we found no differences between service users and a comparison group on reported anxiety after receiving test results, although we did find that this differed by level of comprehension. While this study contributes to understanding the extent and nature of benefits and risks associated with direct lab access, important questions remain for future research about the mechanisms by which these benefits are achieved, how such benefits can be optimized in different health care contexts and for different subgroups of the population, and the specific experiences of patients who receive abnormal test results.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Note on the weight construction.

[PDF File (Adobe PDF File), 127KB - jmir_v17i8e191_app1.pdf]

References

1. Zelmer J, Hagens S. Understanding the gap between desire for and use of consumer health solutions. Healthc Pap 2014;13(4):9-21. [Medline: 25148119]
2. Harris Decima. Annual tracking survey of the general population. Canada Health Infoway 2014.
3. Malone B. Direct Patient Access to Lab Results: Are Labs Ready to Respond to Patient Requests?. Washington, DC: Clinical Laboratory News; 2012. URL: https://www.aacc.org/publications/cln/articles/2012/april/patient-access [accessed 2015-07-27] [WebCite Cache ID 6aKrpu8NF]
4. Giardina D, Singh H. Should patients get direct access to their laboratory test results?: An answer with many questions. JAMA 2011;306(22):2502-2503. [doi: 10.1001/jama.2011.1797]
5. Pyper C, Amery J, Watson M, Crook C. Patients' experiences when accessing their on-line electronic patient records in primary care. Br J Gen Pract 2004 Jan;54(498):38-43 [FREE Full text] [Medline: 14965405]
6. Archer N, Fevrier-Thomas U, Lokker C, McKibbon KA, Straus SE. Personal health records: a scoping review. J Am Med Inform Assoc 2011;18(4):515-522 [FREE Full text] [doi: 10.1136/amiajnl-2011-000105] [Medline: 21672914]
7. Siteman E, Businger A, Gandhi T, Grant R, Poon E, Schnipper J, et al. Clinicians recognize value of patient review of their electronic health record data. In: AMIA Annu Symp Proc. 2006 Presented at: AMIA Annual Symposium proceedings / AMIA Symposium; Nov. 11-15, 2006; Washington, DC p. 1101 URL: http://europepmc.org/abstract/MED/17238720
8. Ammenwerth E, Schnell-Inderst P, Hoerbst A. The impact of electronic patient portals on patient care: a systematic review of controlled trials. J Med Internet Res 2012;14(6):e162 [FREE Full text] [doi: 10.2196/jmir.2238] [Medline: 23183044]
9. Chen C, Garrido T, Chock D, Okawa G, Liang L. The Kaiser Permanente Electronic Health Record: transforming and streamlining modalities of care. Health Aff (Millwood) 2009;28(2):323-333 [FREE Full text] [doi: 10.1377/hlthaff.28.2.323] [Medline: 19275987]

XSL•FO
RenderX

10. Kindler CH, Harms C, Amsler F, Ihde-Scholl T, Scheidegger D. The visual analog scale allows effective measurement of preoperative anxiety and detection of patients' anesthetic concerns. Anesth Analg 2000 Mar;90(3):706-712. [Medline: 10702461]

11. Hassol A, Walker JM, Kidder D, Rokita K, Young D, Pierdon S, et al. Patient experiences and attitudes about access to a patient electronic health care record and linked web messaging. J Am Med Inform Assoc 2004;11(6):505-513 [FREE Full text] [doi: 10.1197/jamia.M1593] [Medline: 15299001]

12. Leveille SG, Walker J, Ralston JD, Ross SE, Elmore JG, Delbanco T. Evaluating the impact of patients' online access to doctors' visit notes: designing and executing the OpenNotes project. BMC Med Inform Decis Mak 2012;12:32 [FREE Full text] [doi: 10.1186/1472-6947-12-32] [Medline: 22500560]

13. Lin C, Wittevrongel L, Moore L, Beaty BL, Ross SE. An Internet-based patient-provider communication system: randomized controlled trial. J Med Internet Res 2005;7(4):e47 [FREE Full text] [doi: 10.2196/jmir.7.4.e47] [Medline: 16236699]

14. Ross SE, Moore LA, Earnest MA, Wittevrongel L, Lin C. Providing a web-based online medical record with electronic communication capabilities to patients with congestive heart failure: randomized trial. J Med Internet Res 2004 May 14;6(2):e12 [FREE Full text] [doi: 10.2196/jmir.6.2.e12] [Medline: 15249261]

15. Giardina T, Menon S, Parrish D. Patient access to medical records and healthcare outcomes: a systematic review. JAMIA 2014;21:737-721. [doi: 10.1136/amiajnl-2013-002239]

Abbreviations

GA-VAS: Global Anxiety-Visual Analog Scale