Journal of Medical Internet Research

Contents

Reviews

XSL•FO
RenderX

Viewpoint

Digital Social Media, Youth, and Nonmedical Use of Prescription Drugs: The Need for Reform

Tim K Mackey[1,2,3], MAS, PhD; Bryan A Liang[1,2,3], MD, JD, PhD; Steffanie A Strathdee[4], PhD

[1]Institute of Health Law Studies, California Western School of Law, San Diego, CA, United States

[2]School of Medicine, Department of Anesthesiology, University of California, San Diego, La Jolla, CA, United States

[3]San Diego Center for Patient Safety - School of Medicine, Department of Anesthesiology, University of California, San Diego, San Diego, CA, United States

[4]Division of Global Public Health, Department of Medicine, University of California, San Diego, San Diego, CA, United States

Corresponding Author:
Tim K Mackey, MAS, PhD
School of Medicine
Department of Anesthesiology
University of California, San Diego
9500 Gilman Drive
La Jolla, CA, 92093-0629
United States
Phone: 1 951 491 4161
Fax: 1 619 515 1599
Email: tmackey@ucsd.edu

Abstract

The tragic death of 18-year-old Ryan Haight highlighted the ethical, public health, and youth patient safety concerns posed by illicit online nonmedical use of prescription drugs (NUPM) sourcing, leading to a federal law in an effort to address this concern. Yet despite the tragedy and resulting law, the NUPM epidemic in the United States has continued to escalate and represents a dangerous and growing trend among youth and adolescents. A critical point of access associated with youth NUPM is the Internet. Internet use among this vulnerable patient group is ubiquitous and includes new, emerging, and rapidly developing technologies—particularly social media networking (eg, Facebook and Twitter). These unregulated technologies may pose a potential risk for enabling youth NUPM behavior. In order to address limitations of current regulations and promote online safety, we advocate for legislative reform to specifically address NUPM promotion via social media and other new online platforms. Using more comprehensive and modernized federal legislation that anticipates future online developments is critical in substantively addressing youth NUPM behavior occurring through the Internet.

KEYWORDS

Introduction

On February 12, 2001, Ryan Haight, an 18-year-old honors student and varsity athlete from California, USA, died from an overdose of the opioid prescription drug Vicodin (hydrocodone/acetaminophen) bought from an online pharmacy without a prescription [1]. His death highlighted the immediate patient safety and public health risks of nonmedical use of prescription medicines (NUPM) by youth (ie, children and adolescents) obtained from the illicit online environment. This tragic event led to passage of the 2008 US federal legislation, the Ryan Haight Online Pharmacy Consumer Protection Act (RHA), which established regulatory provisions and tools for the Drug Enforcement Agency (DEA) of the United States to control the sale and dispensing of controlled substances over the Internet [2].

However, the effectiveness of the RHA on NUPM online sourcing and regulation of online pharmacies has not been well established or studied. Consequently, the problem of illicit online sourcing of controlled substances and other medications without a prescription has yet to be adequately resolved [1,3]. Despite RHA passage, new forms of unregulated digital media

and information technology platforms continue to be developed and are rapidly becoming associated with illicit online prescription drug sourcing in digital environments highly populated by youth.

In order to inform policy efforts to address youth NUPM and current regulatory limitations, we explore the potential public health and patient safety implications of promotion of youth-based NUPM in social media. To do so, we first review current national trends in youth NUPM behavior and Internet and social media utilization. We then examine the use of social media by illicit online pharmacies in promoting NUPM and analyze current policy instruments, including the RHA. We then recommend policy solutions and advocate for additional research to better inform the public and ensure safe Internet access to prevent youth NUPM.

Nonmedical Use of Prescription Drugs

National Trends in NUPM

Since Ryan Haight's death, prescription drug abuse among youth has become part of a larger national trend of morbidity and mortality associated with drug overdose, diversion, and polydrug abuse [4-7]. The US Centers for Disease Control and Prevention (CDC) reported in 2010 that more than 12 million people engaged in nonmedical use of prescription painkillers alone; misuse/abuse of this drug class was responsible for approximately 475,000 emergency room admissions in 2009 [6,8]. Indeed, misuse has led to a marked increase in US public and private health care expenditures, estimated up to $72.5 billion in direct costs annually [8-10]. Prescription drug abuse also disproportionately impacts vulnerable populations, including rural groups, low-income groups, those subject to sexual victimization or dating violence, those with a history of mental illness, and those with a history of substance abuse disorders [5,6,11-13].

Youth NUPM

Crucially, a key high-risk group for NUPM is youth (children and adolescents, aged 12-17). Estimated prevalence of NUPM among this age group is high, with the CDC reporting in 2011 that 20.7% of high school students had engaged in NUPM (OxyContin, Percocet, Vicodin, Adderall, Ritalin or Xanax) [14]. A 2010 National Survey on Drug Use and Health similarly reported at least 3.0% of all youths (and 5.9% of 18-25 year olds) reported psychotherapeutic NUPM in the past month in 2010 [4,15]. Other studies report even higher prevalence of abuse [12,16]. More recently, the Monitoring the Future 2011 national survey reported that after marijuana, prescription and over-the-counter medications represented the most commonly abused drugs for either licit or illicit drugs among 12th graders [17]. Most troubling, NUPM use among youth may also lead to other forms of substance and illicit drug abuse [4,7,18-20].

Youth narcotic and controlled substance NUPM, which includes commonly abused pain medications such as OxyContin and Vicodin, is perhaps the most deeply concerning risk to youth health development and has even been associated with illicit heroin drug abuse [1,21]. NUPM in these drug classes can result in severe adverse clinical interactions and side effects, drug

dependency, and increased emergency room admissions [2,10,16,22]. NUPM is also associated with other high-risk health behavior including alcohol consumption and marijuana use, resulting in poor school performance—yet may be perceived as a lower risk behavior by youth compared to other forms of illicit drug abuse [1,3,12,15,17]. Additionally, increases in NUPM associated with Attention Deficit Hyperactivity Disorder drugs (ADHD), such as Ritalin and Adderall, have become a serious concern [22-25]. Importantly, virtually all these drugs have been detected as marketed by illicit "no prescription" online pharmacies and have been subject to counterfeiting [1,2,25-28].

Traditional NUPM Sourcing

Traditional methods of drug diversion, including person-to-person purchasing, trading, loaning, sharing, stealing and theft, family member and friend access, street drug purchases, prescription forgeries and fraud, and "doctor/prescription shopping", have traditionally enabled NUPM [1,12,16]. In order to address these vulnerabilities, some US states have implemented "Prescription Drug Monitoring Programs" (PDMPs) to track prescribing and dispensing of controlled substances in order to detect suspected abuse and diversion [4-7,16].

Although these programs may provide controls to stem diversion of high-risk prescription drugs to youth populations [6,8,16], they are highly uneven in enforcement and state resource commitment [29]. Consequently, they may be ineffective for broader identification and intercession in youth NUPM sourcing. But further, uneven PDMPs may not be responsive to the changing nature of health information seeking and online behavior associated with youth NUPM. Indeed, PDMPs may miss the mark in terms of where youth NUPM sourcing is starting to occur and do not address online sourcing of prescription drugs, which may be familiar to youth yet difficult to trace for illicit activity [22,30-32]. To date, this specific risk factor has not been adequately assessed in youth-related NUPM research.

Potential Online Risks for NUPM Behavior

Internet and Social Media Utilization Trends

Exacerbating risks of NUPM access is unregulated content on the Internet, the use of which is now ubiquitous among both youth and adults. Indeed, survey data from the Pew Research Center's Internet and American Life Project (Pew Internet) indicate that some 72% of US adult Internet users search for health and medical information online and that more than one third engage in health care self-diagnosing [33,34]. In addition, the US Food and Drug Administration (FDA) reports 23% of adult Internet consumers have admitted to purchasing a prescription medicine online, of whom 15% acknowledged the risky nature of purchasing from an online pharmacy located outside the United States [35].

As might be expected, Internet use by the youth demographic is almost universal. Pew Internet reports that an estimated 95% of teens (ages 12-17) [36] currently use the Internet and are the most likely age groups to have an online presence [37]. In

addition, there has been a rapid rise in utilization of social networking reflected by a majority (80%) of online teens using popular social media sites including Facebook (93%), and use of other social media platforms including Myspace (24%), Twitter (12%), and YouTube (6%) [36]. Indeed, youth respondents have reported that the Internet is their primary source of general information, even if the credibility of such information is difficult to determine [38].

Although this population group has widespread adoption of the Internet and social media, they may not engage in safe online behavior. For example, at least 44% of teens admitted they lie about their age to access websites or to set up an online account [36]. Indeed, those using social media sites report being twice as likely as nonusers to misrepresent their age [36]. At the same time, teens are reporting that they use online sources for looking up health, dieting, and physical fitness information (31%) and that 17% of them go online for information on difficult topics including drug use and sexual health [37].

Within this already vulnerable population, there is also a disproportionate income effect. Teens from lowest-income families are twice as likely (23% vs 11%) to seek health information online compared to teens from higher income households [37]. Further, almost half (48%) of teens report purchasing items online, indicating that teens may be comfortable and have access to make potentially illicit purchases if appropriate controls are absent [37].

NUPM and Illicit Online Pharmacies

Youth online behavior trends indicate that this population is adopting digital technology for consumption of health information and may be engaged in risky online behavior, which can increase risk for Internet-enabled NUPM [37,39]. Several studies have identified the public health risks of sourcing from "no prescription" illicit online pharmacies that enable NUPM, including among youth and adolescents [9,22,26,27,30,32,40,41].

Importantly, any online pharmacy purportedly marketing the sale of a prescription medication without the need of a prescription is both violating applicable US laws and regulations, as well as promoting NUPM behavior given that adequate controls to ensure patient safety are lacking. This promotion of NUPM is often facilitated by false and misleading marketing used in online direct-to-consumer advertising (DTCA) [42,43], which has yet to be adequately regulated by FDA and others [42,43]. These illicit forms of DTCA may be difficult for consumers, particularly youth, to accurately identify as legitimate (or not), despite public service announcements attempting to inform consumers that online purchasing can be dangerous [44].

Despite its illegality, the spectrum of drugs available for online NUPM sourcing is virtually unlimited [1]. This includes a host of therapeutic drug classes marketed without sufficient controls, including drugs for weight loss, ADHD, steroids, inhalants, contraception drugs and devices, opioids, a variety of narcotics, and drugs in critical shortage promoted across various Internet mediums, including social media [9,22,24,26,30,32,40,41,43,45,46].

Collectively, these studies illustrate that illicit online sourcing represents a potential risk factor for youth NUPM. Illicit NUPM promotion through Internet pharmacies engenders a completely unregulated system of parallel access for youth. This can lead to self-prescribing of virtually any medicine, resulting in drug abuse and dependence, as well as use of drug forms that are of questionable quality, authenticity, and safety, all without medical or parental oversight [1,28]. Tragically, this form of NUPM sourcing has been directly linked to patient deaths, including Ryan Haight, as well as others [1,28].

Lack of Sufficient Research on Social Media and NUPM

Recognition and needed research on the convergence of social media and youth NUPM is highly uneven. Despite growing evidence of online sourcing risks, a recent systematic review of NUPM behavior among adolescents failed to mention online information seeking/sourcing or social media usage as a specific risk factor [12]. Conversely, organizations such as the National Center on Addiction and Substance Abuse have *specifically* identified increased risks associated with substance abuse for youth who use social media [39]. The United Nations International Narcotics Control Board also warns that illicit Internet pharmacies have started using social media to target young audiences [47].

Some studies have also attempted to assess this area of risk. Previous research has identified increasing use of popular social media platforms by illicit "no prescription" online pharmacies marketing the sale of several high-risk drug products [24,43,46,48]. This includes a recent study that found that illegal DTCA marketing of a fictitious illicit online pharmacy using social media sites Facebook, Myspace, and Twitter was easily accessible and could be done at low cost [48].

Another published study examined the use of Twitter to discuss Adderall NUPM behavior among college students [49]. It found that 8.9% of Adderall-related tweets analyzed mentioned another substance (including illicit drugs), indicating the dangerous possibility of promotion of polydrug abuse via social media [49]. Another unpublished study analyzed Adderall-related Twitter traffic and found that the highest volume of Twitter content (roughly 7 out of 10) originated from illicit online pharmacies advertising the sale of medications with no prescription required [50].

Though an evidence base supporting the association between social media and NUPM is beginning to emerge, there is an urgent need for additional research specifically examining in detail NUPM-related risk factors enabled by social media. This should be pursued in conjunction with policy analysis to determine if current law and legislation can effectively regulate this digital medium to ensure youth and patient safety.

Ineffective Enforcement/Coverage of Existing Regulations

More than 10 years after Ryan Haight's unintentional death, youth online-enabled NUPM access remains relatively unabated despite legislative and law enforcement efforts. Global action (such as Interpol's Operation Pangea) have led to the closure of some illicit online pharmacies [51]. Yet despite these

operations, organizations such as the National Association of Boards of Pharmacy (NABP) continue to report that the vast majority (97%) of existing online pharmacies are "not recommended" and present potential patient safety risks [52]. This includes 87% of recent NABP-reviewed online pharmacies not requiring a valid prescription for dispensing [52].

The specific mechanisms of the RHA to stem controlled substance online NUPM focus on registration, licensure, disclosure, and reporting requirements for online pharmacies offering controlled substances as well as requiring valid prescriptions for dispensing (including at least one in-person examination) [2]. It also imposes increased penalties for illicit actors in an attempt to deter such criminal activity [2]. Yet, the RHA primarily focuses on domestic online pharmacies, which is problematic given that surveys have identified up to 23% having a physical addresses outside the United States and most do not provide any address at all [1,52]. Hence, online pharmacies selling controlled substances that operate outside of the United States may not be subject to the jurisdiction of the Act or the DEA, limiting enforceability.

Additional gaps in the RHA in effectively dealing with the illicit online sale of controlled substances have also been reported. This includes websites "unlocking" hidden content that provides access to controlled substances and using affiliate networks and portal sites to avoid law enforcement detection [3]. Further, other illicit actors may simply sell the "prescription" to the patient for an additional fee, allowing for re-use and may not be subject to the RHA (1). Criminals operating online pharmacies have also gone as far as impersonating DEA agents and defrauding consumers by threatening law enforcement and prosecution for illegal purchase of a drug after a consumer has purchased online [53].

Further highlighting the limitations of the RHA in effectively regulating controlled substance NUPM, a 2011 report by online monitoring company LegitScript, published a sample list of 1000 illicit online pharmacies actively offering the sale of controlled substances without a valid prescription (including over half with domain name or server presence in the United States)—an activity in direct violation of the RHA [3]. Yet, despite these clear legal violations and claims by DEA of RHA effective deterrence, there appears to be little enforcement with no successful prosecutions under the RHA against these or other criminal violations of the law [3,54]. Hence, there is a clear need to reexamine the scope and coverage of the RHA and enable additional tools of enforcement to meet changing online trends and current regulatory gaps.

Reform

Amending the Ryan Haight Act

Associated risks of NUPM to patient safety and public health are high, but in no group is the risk greater than in youth and adolescents. The physical, mental health, and emotional harms from youth NUPM can have lasting impacts for this vulnerable population [17]. Yet, the combination of the continuing national public health crisis of youth NUPM, increased utilization by youth of the Internet and social media, and an insufficiently

regulated online environment that allows NUPM promotion and sourcing continue to put youth at significant risk. Though illicit online pharmacies that enable NUPM behavior present a global public health problem requiring international cooperation, even at the domestic level, amendments to the existing RHA could improve effectiveness and enforceability to better prevent youth NUPM.

Reform should begin with examining amendment and modernization of the RHA to improve its scope, effectiveness, and enforceability over illicit online NUPM promotion of prescription controlled substance drugs where it is actively occurring. First, the RHA does not specifically address other nonInternet pharmacy actors that actively facilitate this illicit trade. These Internet service intermediaries are clearly enabling NUPM behavior and sourcing and may also profit from this illicit activity through generation of revenue from search engine marketing/optimization, ad revenue, and processing, membership, and referral fees [1,43]. Specifically, the RHA does not address NUPM promotion through social media, though these forums have already been identified as allowing promotion of NUPM by illicit online pharmacies [43,48,49].

These enabling risk factors require RHA amendment to expand its scope and enforcement powers to address new forms of digital communication and media that promote online NUPM. This could be accomplished by amending the RHA to include a new definition of "Enabling 3rd Party Intermediaries" to capture additional and relevant online mediums promoting NUPM and illicit access points. Through amending the RHA to include this provision, this term can encompass online digital technologies, including nonpharmacy websites, Internet service providers (ISPs), Web applications, mobile-based platforms/games, payment processors, affiliate sites, membership forums, and, specifically, social media sites. Further, it can focus on high-risk and clearly illegal online promotion activities that advertise sourcing without a prescription, facilitate NUPM sourcing (through direct links to online pharmacies, online ads, etc), or fail to monitor and remove direct marketing associated with NUPM promotion often in direct violation with their own legal terms of conditions and use (including key social media platforms) [48].

Indeed, despite potential facilitation of illicit sourcing, third-party sites have remained largely unregulated and have for the most part escaped enforcement efforts [28]. One clear exception has been the world's largest search engine Google, which was fined $500 million by the US Department of Justice in 2011 for illegal online pharmacy ads that led to a change in its AdWords program [55]. Hence, by pursuing amendment of the RHA, if any of these third-party intermediaries in fact have a physical location or infrastructure in the United States, jurisdiction could be extended over them and their actions could be made subject to the enforcement provisions of the Act [43]. In this way, the entire digital ecosystem of online-enabled NUPM can be addressed through simple amendment of existing legislation enabling the DEA to pursue more proactive enforcement actions to promote public health.

In addition, in order to provide consumers with important and necessary information on safe online sourcing of controlled

substances, the RHA should also be amended to require the DEA to publish a publicly available list of online pharmacies that have successfully modified their DEA registration to allow online sale of controlled substances as statutorily required under the Act [2]. This list of authorized and registered DEA online pharmacies should also incorporate with NABP verification through its Verified Internet Pharmacy Practice Sites (VIPPS) program, the only system recommended by the FDA. This would better ensure ongoing RHA compliance and state licensure verification and better inform consumers about safe online sourcing. Additionally, use of monitoring companies such as LegitScript, which has clients such as Google, Microsoft, Amazon, and the FDA, may better ensure that authorized sites are actively monitored and remain compliant with RHA mandates.

Lessons From Past Legislation

Past failed legislative efforts to more dynamically regulate online pharmacies may provide important lessons for future potential solutions. In 2012, two US congressional bills, the House's Stop Online Piracy Act and the Senate's Protect IP Act, included provisions to regulate domestic and foreign online pharmacy websites, and associated search engines, payment processors, and other ISPs, that facilitate illicit online drug e-commerce [56]. However, these bills also contained additional intellectual property rights provisions for other forms of digital medium and online services (eg, videos, music, etc) not related to health that became the subject of controversy and protest and led to the defeat of both bills [56].

In retrospect, it seems clear that important public health considerations to protect consumers online need to be positioned in their own unique legislation that solely addresses issues of patient safety and does not concurrently address commercial or intellectual property rights. Hence, an amendment of the RHA may provide for such a policy forum, as controlled substance NUPM among youth and adolescents continues to represent a national public health crisis that existing law has arguably failed to adequately address and the subject is sufficiently narrow in scope compared to general anticounterfeiting legislation. Though legislative action may face challenges, amendment of the RHA could modernize the Act to respond to emerging digital technologies, provide additional tools to the DEA in pursuing enforcement, and address regulatory gaps currently being exploited by illicit online pharmacies

Increased Policy Advocacy and Action

Last, there is a need for better cooperation and tangible action by stakeholders currently advocating for action against illicit online pharmacies. The Center for Safe Internet Pharmacies (CSIP), a nonprofit organization with goals of combating illegal online pharmacies through education, enforcement, and information dissemination, was formed in 2011 and is a partnership between numerous private sector entities actively involved in e-commerce [57].

Included in CSIP membership as a strategic partner and board member is the world's largest social media platform, Facebook [58]. Yet, despite its membership and apparent public engagement on this issue, recent research indicates that social media sites do little to enforce their own policies or monitor their content for NUPM-related promotion [48]. CSIP and other organizations, such as the Association of Safe Online Pharmacies, must more actively engage member ISPs and other stakeholders they partner with to prioritize accountability and enforcement against clearly illicit NUPM promotion, especially that which targets youth.

Conclusions

The frenetic pace of technology change through new forms of digital sources has quickly made existing legislative approaches to maintain online drug safety antiquated. This is reflected in today's "Ryan Haight", who is not only frequently on the Internet but is also a common if not daily user of popular social media sites such as Facebook, a platform already linked to NUPM promotion [24,43,48]. He or she may also be in any part of the world with Internet access and subsequently has access to a global illicit online trade of suspect medicines that bypasses country borders and rule of law.

Hence, it is crucial that particularly youth, who are already at high risk of NUPM and are the most active demographic on the Internet, be provided a safe online environment to make rational and informed choices not to engage in dangerous health behavior. Unfortunately, the present environment presents significant challenges for this important decision-making process and attempts at prevention. Hence, domestic and international approaches addressing NUPM must be modernized to meet the needs of a new digital youth generation and prevent the unnecessary death of the next Ryan Haight.

or activities that could appear to have influenced the submitted work. Steffanie A. Strathdee reports no conflicts of interest associated with this manuscript.

References

1. Liang BA, Mackey T. Searching for safety: addressing search engine, website, and provider accountability for illicit online drug sales. Am J Law Med 2009;35(1):125-184. [Medline: 19534258]
2. Department of Justice. Implementation of the Ryan Haight Online Pharmacy Consumer Protection Act of 2008. URL: http://www.deadiversion.usdoj.gov/fed_regs/rules/2009/fr0406.pdf [accessed 2012-11-30] [WebCite Cache ID 6CYOm4Xvf]
3. LegitScript. Drug Dealers on the Internet: Is the DEA enforcing the Ryan Haight Act?. 2011. URL: http://www.legitscript.com/download/LegitScript-DEA-Rogue-Internet-Pharmacy-Analysis.pdf [accessed 2013-05-21] [WebCite Cache ID 6GmvLwQFc]
4. Boyd CJ, Young A, Grey M, McCabe SE. Adolescents' nonmedical use of prescription medications and other problem behaviors. J Adolesc Health 2009 Dec;45(6):543-550 [FREE Full text] [doi: 10.1016/j.jadohealth.2009.03.023] [Medline: 19931825]
5. Centers for Disease Control and Prevention (CDC). CDC grand rounds: prescription drug overdoses - a U.S. epidemic. MMWR Morb Mortal Wkly Rep 2012 Jan 13;61(1):10-13 [FREE Full text] [Medline: 22237030]
6. Centers for Disease Control and Prevention (CDC). Policy Impact: Prescription Painkiller Overdoses. URL: http://www.cdc.gov/homeandrecreationalsafety/rxbrief/ [accessed 2012-11-30] [WebCite Cache ID 6CYOvN9cC]
7. Hernandez SH, Nelson LS. Prescription drug abuse: insight into the epidemic. Clin Pharmacol Ther 2010 Sep;88(3):307-317. [doi: 10.1038/clpt.2010.154] [Medline: 20686478]
8. Centers for Disease Control and Prevention (CDC). Prescription Painkiller Overdoses in the US. URL: http://www.cdc.gov/vitalsigns/painkilleroverdoses/ [accessed 2012-11-30] [WebCite Cache ID 6CYOxkqvM]
9. Jena AB, Goldman DP. Growing Internet use may help explain the rise in prescription drug abuse in the United States. Health Aff (Millwood) 2011 Jun;30(6):1192-1199. [doi: 10.1377/hlthaff.2011.0155] [Medline: 21565838]
10. Wu LT, Pilowsky DJ, Patkar AA. Non-prescribed use of pain relievers among adolescents in the United States. Drug Alcohol Depend 2008 Apr 1;94(1-3):1-11 [FREE Full text] [doi: 10.1016/j.drugalcdep.2007.09.023] [Medline: 18054444]
11. Young A, Grey M, Boyd CJ, McCabe SE. Adolescent Sexual Assault and the Medical and Nonmedical Use of Prescription Medication. J Addict Nurs 2011 Mar;11(1-2):25-31 [FREE Full text] [doi: 10.3109/10884601003628138] [Medline: 22065397]
12. Young AM, Glover N, Havens JR. Nonmedical use of prescription medications among adolescents in the United States: a systematic review. J Adolesc Health 2012 Jul;51(1):6-17. [doi: 10.1016/j.jadohealth.2012.01.011] [Medline: 22727071]
13. Temple JR, Freeman DH. Dating violence and substance use among ethnically diverse adolescents. J Interpers Violence 2011 Mar;26(4):701-718. [doi: 10.1177/0886260510365858] [Medline: 20587475]
14. Eaton DK, Kann L, Kinchen S, Shanklin S, Flint KH, Hawkins J, Centers for Disease Control and Prevention (CDC). Youth risk behavior surveillance - United States, 2011. MMWR Surveill Summ 2012 Jun 8;61(4):1-162 [FREE Full text] [Medline: 22673000]
15. SAMHSA. Results from the 2010 National Survey on Drug Use and Health: Summary of National Findings. URL: http://www.samhsa.gov/data/NSDUH/2k10NSDUH/2k10Results.htm#Ch8 [accessed 2012-11-30] [WebCite Cache ID 6CYP0U7Lr]
16. Fischer B, Bibby M, Bouchard M. The global diversion of pharmaceutical drugs non-medical use and diversion of psychotropic prescription drugs in North America: a review of sourcing routes and control measures. Addiction 2010 Dec;105(12):2062-2070. [doi: 10.1111/j.1360-0443.2010.03092.x] [Medline: 20840172]
17. Johnston LD, O'Malley PM, Bachman JG, Schulenberg JE. Institute for Social Research. Monitoring the Future National Results on Adolescent Drug Use: Overview of Key Findings, 2011 URL: http://monitoringthefuture.org/pubs/monographs/mtf-overview2011.pdf [accessed 2012-11-30] [WebCite Cache ID 6CYP2mayR]
18. Catalano RF, White HR, Fleming CB, Haggerty KP. Is nonmedical prescription opiate use a unique form of illicit drug use? Addict Behav 2011 Jan;36(1-2):79-86 [FREE Full text] [doi: 10.1016/j.addbeh.2010.08.028] [Medline: 20864261]
19. Sung HE, Richter L, Vaughan R, Johnson PB, Thom B. Nonmedical use of prescription opioids among teenagers in the United States: trends and correlates. J Adolesc Health 2005 Jul;37(1):44-51. [doi: 10.1016/j.jadohealth.2005.02.013] [Medline: 15963906]
20. McCabe SE, Teter CJ, Boyd CJ. Medical use, illicit use, and diversion of abusable prescription drugs. J Am Coll Health 2006;54(5):269-278 [FREE Full text] [doi: 10.3200/JACH.54.5.269-278] [Medline: 16539219]
21. Siegal HA, Carlson RG, Kenne DR, Swora MG. Probable relationship between opioid abuse and heroin use. Am Fam Physician 2003 Mar 1;67(5):942, 945 [FREE Full text] [Medline: 12643356]
22. Schepis TS, Marlowe DB, Forman RF. The availability and portrayal of stimulants over the Internet. J Adolesc Health 2008 May;42(5):458-465 [FREE Full text] [doi: 10.1016/j.jadohealth.2007.11.140] [Medline: 18407040]
23. Kaye S, Darke S. The diversion and misuse of pharmaceutical stimulants: what do we know and why should we care? Addiction 2012 Mar;107(3):467-477. [doi: 10.1111/j.1360-0443.2011.03720.x] [Medline: 22313101]

24. Liang BA, Mackey TK. Online availability and safety of drugs in shortage: a descriptive study of internet vendor characteristics. J Med Internet Res 2012;14(1):e27 [FREE Full text] [doi: 10.2196/jmir.1999] [Medline: 22321731]

25. Yukhananov A. Fake ADHD drug Adderall sold online, U.&#46;S.&#46; FDA warns.: Reuters; 2012. URL: http://www.reuters.com/article/2012/05/29/us-drugs-fda-adderall-idUSBRE84S1FO20120529 [accessed 2012-11-30] [WebCite Cache ID 6CYP620jR]

26. Forman RF. Availability of opioids on the Internet. JAMA 2003 Aug 20;290(7):889. [doi: 10.1001/jama.290.7.889] [Medline: 12928464]

27. Forman RF, Block LG. The marketing of opioid medications without prescription over the Internet. Journal of Public Policy & Marketing 2006 Sep;25(2):133-146. [doi: 10.1509/jppm.25.2.133]

28. Mackey TK, Liang BA. The global counterfeit drug trade: patient safety and public health risks. J Pharm Sci 2011 Nov;100(11):4571-4579. [doi: 10.1002/jps.22679] [Medline: 21698604]

29. Bureau of Justice Assistance. What Have We Learned From Evaluations of PDMP Programs?. URL: https://www.bja.gov/evaluation/program-substance-abuse/pdmp2.htm [accessed 2012-11-30] [WebCite Cache ID 6CYPNtGte]

30. Forman RF, Woody GE, McLellan T, Lynch KG. The availability of web sites offering to sell opioid medications without prescriptions. Am J Psychiatry 2006 Jul;163(7):1233-1238. [doi: 10.1176/appi.ajp.163.7.1233] [Medline: 16816229]

31. Nielsen S, Barratt MJ. Prescription drug misuse: is technology friend or foe? Drug Alcohol Rev 2009 Jan;28(1):81-86. [doi: 10.1111/j.1465-3362.2008.00004.x] [Medline: 19320680]

32. Klein CA, Kandel S. www.mydrugdealer.com: Ethics and legal implications of Internet-based access to substances of abuse. J Am Acad Psychiatry Law 2011;39(3):407-411. [Medline: 21908759]

33. Fox S, Duggan M. For one-third of U.&#46;S.&#46; adults, the internet is a diagnostic tool.: Pew Internet Research URL: http://pewinternet.org/Reports/2013/Health-online/Part-One/Section-1.aspx [accessed 2013-04-20] [WebCite Cache ID 6G29kmfbQ]

34. Kuehn BM. More than one-third of US individuals use the Internet to self-diagnose. JAMA 2013 Feb 27;309(8):756-757. [doi: 10.1001/jama.2013.629] [Medline: 23443421]

35. FDA. BeSafeRx: Know Your Online Pharmacy - Survey Highlights. URL: http://www.fda.gov/Drugs/ResourcesForYou/Consumers/BuyingUsingMedicineSafely/BuyingMedicinesOvertheInternet/BeSafeRxKnowYourOnlinePharmacy/ucm318497.htm [accessed 2013-04-20] [WebCite Cache ID 6G29z6Gju]

36. Brenner J. Pew Internet: Teens. URL: http://pewinternet.org/Commentary/2012/April/Pew-Internet-Teens.aspx [accessed 2012-11-30] [WebCite Cache ID 6CYPCdfFR]

37. Lenhard A, Purcell K, Smith A, Zickuhr K. Social Media &#38; Mobile Internet Use Among Teens and Young Adults. URL: http://pewinternet.org/~/media//Files/Reports/2010/PIP_Social_Media_and_Young_Adults_Report_Final_with_toplines.pdf [accessed 2012-11-30] [WebCite Cache ID 6CYPA4ylh]

38. Gray NJ, Klein JD, Noyce PR, Sesselberg TS, Cantrill JA. Health information-seeking behaviour in adolescence: the place of the internet. Soc Sci Med 2005 Apr;60(7):1467-1478. [doi: 10.1016/j.socscimed.2004.08.010] [Medline: 15652680]

39. National Center on Addiction and Substance Abuse at Columbia University. National Survey of American Attitudes on Substance Abuse XVI: Teens and Parents. URL: http://www.casacolumbia.org/templates/PressReleases.aspx?articleid=650&zoneid=87 [accessed 2012-11-30] [WebCite Cache ID 6CYPEtOJe]

40. Sikes A, Walley C, McBride R, Fusco A, Cole RF, Lauka J. Inhalant and Prescription Medication Abuse among Adolescents: An Inexpensive, Accessible, and Misperceived Trend. Journal of Child & Adolescent Substance Abuse 2011 Jul;20(3):237-252. [doi: 10.1080/1067828X.2011.581903]

41. Clement CL, Marlowe DB, Patapis NS, Festinger DS, Forman RF. Nonprescription steroids on the Internet. Subst Use Misuse 2012 Feb;47(3):329-341 [FREE Full text] [doi: 10.3109/10826084.2011.630225] [Medline: 22080724]

42. Liang BA, Mackey T. Direct-to-consumer advertising with interactive internet media: global regulation and public health issues. JAMA 2011 Feb 23;305(8):824-825. [doi: 10.1001/jama.2011.203] [Medline: 21343583]

43. Liang BA, Mackey TK. Prevalence and Global Health implications of social media in direct-to-consumer drug advertising. J Med Internet Res 2011;13(3):e64 [FREE Full text] [doi: 10.2196/jmir.1775] [Medline: 21880574]

44. Ivanitskaya L, Brookins-Fisher J, O Boyle I, Vibbert D, Erofeev D, Fulton L. Dirt cheap and without prescription: how susceptible are young US consumers to purchasing drugs from rogue internet pharmacies? J Med Internet Res 2010;12(2):e11 [FREE Full text] [doi: 10.2196/jmir.1520] [Medline: 20439253]

45. Friedman RA. The changing face of teenage drug abuse--the trend toward prescription drugs. N Engl J Med 2006 Apr 6;354(14):1448-1450. [doi: 10.1056/NEJMp068010] [Medline: 16598042]

46. Liang BA, Mackey TK, Lovett KM. Suspect online sellers and contraceptive access. Contraception 2012 Nov;86(5):551-556. [doi: 10.1016/j.contraception.2012.04.011] [Medline: 22633248]

47. United Nations Office on Drugs and Crime. International Narcotics Control Board warns of illegal online pharmacies selling drugs to youth. URL: http://www.unodc.org/unodc/en/frontpage/2012/February/narcotics-control-board-warns-of-illegal-online-pharmacies-selling-drugs-to-youth.html [accessed 2012-11-30] [WebCite Cache ID 6CYPTAB3r]

48. Mackey TK, Liang BA. Global Reach of Direct-to-Consumer Advertising Using Social Media for Illicit Online Drug Sales. J Med Internet Res 2013;15(5):e105 [FREE Full text] [doi: 10.2196/jmir.2610] [Medline: 23718965]

49. Hanson CL, Burton SH, Giraud-Carrier C, West JH, Barnes MD, Hansen B. Tweaking and tweeting: exploring Twitter for nonmedical use of a psychostimulant drug (Adderall) among college students. J Med Internet Res 2013;15(4):e62 [FREE Full text] [doi: 10.2196/jmir.2503] [Medline: 23594933]

50. Kemp C. Mid-term Prescription Drug Abuse: Does Social Media Play a Role?. URL: http://cathykemp.blogspot.com/2012/03/mid-term-prescription-drug-abuse-does.html [accessed 2013-05-21] [WebCite Cache ID 6Gmvxslmw]

51. Interpol. Operations. URL: http://www.interpol.int/Crime-areas/Pharmaceutical-crime/Operations/Operation-Pangea [accessed 2013-05-21] [WebCite Cache ID 6GmvuysIj]

52. National Association of Boards of Pharmacy. Buying Medicine Online. URL: http://www.nabp.net/programs/consumer-protection/buying-medicine-online/ [accessed 2012-11-30] [WebCite Cache ID 6CYPbCo0I]

53. Harris D. Prescription Drug Scam Uses DEA Agents&#39; Identities To Terrify Victims into Paying Up. URL: http://abcnews.go.com/US/prescription-drug-scam-dea-agents-identities-terrify-victims/story?id=16116808 [accessed 2013-05-21] [WebCite Cache ID 6Gmvt2q9H]

54. Ellis L. Let 1,000 Online Pharmacies Bloom?. 2011. URL: http://www.motherjones.com/blue-marble/2011/06/should-online-pharma-business-be-shut-down [accessed 2013-05-21] [WebCite Cache ID 6Gmvnm6e3]

55. Department of Justice. Google Forfeits $500 Million Generated by Online Ads &#38; Prescription Drug Sales by Canadian Online Pharmacies. URL: http://www.justice.gov/opa/pr/2011/August/11-dag-1078.html [accessed 2013-05-21] [WebCite Cache ID 6GmvmBVHS]

56. Mackey TK, Liang BA. Separating online piracy from counterfeit medicines: The need for policy reform and a call to action. Health Policy and Technology 2012 Sep;1(3):120-122. [doi: 10.1016/j.hlpt.2012.07.004]

57. The Center for Safe Internet Pharmacies. Who We Are. URL: http://www.safemedsonline.org/who-we-are/ [accessed 2013-05-21] [WebCite Cache ID 6Gmvfpv3s]

58. The Center for Safe Internet Pharmacies. Members. URL: http://www.safemedsonline.org/who-we-are/members/ [accessed 2013-05-21] [WebCite Cache ID 6GmvgnfMd]

Abbreviations

ADHD: Attention Deficit Hyperactivity Disorder
CDC: US Centers for Disease Control and Prevention
CSIP: Center for Safe Internet Pharmacies
DEA: US Drug Enforcement Agency
DTCA: direct-to-consumer advertising
FDA: US Food and Drug Administration
ISPs: Internet Service Providers
NABP: National Association of Boards of Pharmacy
NUPM: nonmedical use of prescription medicines
PDMPs: Prescription Drug Monitoring Programs
Pew Internet: Pew Research Center's Internet and American Life Project
RHA: Ryan Haight Online Pharmacy Consumer Protection Act
VIPPS: Verified Internet Pharmacy Practice Sites

XSL•FO
RenderX

Viewpoint

Principles and Framework for eHealth Strategy Development

Richard E Scott[1,2], BSc (Hons), PhD; Maurice Mars[1], MB ChB, MD

[1]Nelson R Mandela School of Medicine, Department of TeleHealth, University of KwaZulu-Natal, Durban, South Africa
[2]Office of Global e-Health and Strategy, Faculty of Medicine, University of Calgary, Calgary, AB, Canada

Corresponding Author:
Richard E Scott, BSc (Hons), PhD
Nelson R Mandela School of Medicine
Department of TeleHealth
University of KwaZulu-Natal
Private Bag 7
Congella 4013
Durban,
South Africa
Phone: 27 31 260 4543
Fax: 27 31 260 4737
Email: rscott@ukzn.ac.za

Abstract

Significant investment in eHealth solutions is being made in nearly every country of the world. How do we know that these investments and the foregone opportunity costs are the correct ones? Absent, poor, or vague eHealth strategy is a significant barrier to effective investment in, and implementation of, sustainable eHealth solutions and establishment of an eHealth favorable policy environment. Strategy is the driving force, the first essential ingredient, that can place countries in charge of their own eHealth destiny and inform them of the policy necessary to achieve it. In the last 2 years, there has been renewed interest in eHealth strategy from the World Health Organization (WHO), International Telecommunications Union (ITU), Pan American Health Organization (PAHO), the African Union, and the Commonwealth; yet overall, the literature lacks clear guidance to inform countries why and how to develop their own complementary but locally specific eHealth strategy. To address this gap, this paper further develops an eHealth Strategy Development Framework, basing it upon a conceptual framework and relevant theories of strategy and complex system analysis available from the literature. We present here the rationale, theories, and final eHealth strategy development framework by which a systematic and methodical approach can be applied by institutions, subnational regions, and countries to create holistic, needs- and evidence-based, and defensible eHealth strategy and to ensure wise investment in eHealth.

Introduction

Many definitions of eHealth have been developed or adopted, but perhaps the bottom-line message is that eHealth can be anything we want it to be. It is simply the application of information and communications technologies (ICTs) to the health sector [1]. Evidence shows eHealth is now a globally pervasive tool [2] yet seldom have health organizations, countries, or geographic regions had a proper eHealth strategy to guide implementation. Why then the renewed discussion about "eHealth strategy" in developing countries and regions during 2011? For example, the Pan American Health Organization (PAHO) promulgated its Regional eHealth

Strategy approach at their 51st Directing Council meeting [3], and in February 2011 the African Union resurrected past debate around the issue at an Experts Meeting on eHealth and Telemedicine Harmonization in Africa. Similarly, Kenya just completed a 2-year undertaking to develop its eHealth strategy [4], and South Africa has just released its revised eHealth Strategy [5]. In 2012, a WHO/ITU collaboration released its WHO-ITU National eHealth Strategy Toolkit [6]. Perhaps the need is finally being understood. However, although these documents provide some insight, specific guidance for individual countries or institutions to design and develop their own eHealth strategy is unclear and is lacking in the literature.

XSL•FO
RenderX

As a consequence, entities will often emulate or adapt practice from elsewhere. While emulation or adaptation is common, these approaches are inappropriate: "emulation" because solutions and approaches must be context specific, and "adaptation" because, although a compromise, it remains suboptimal. A sustainable eHealth solution is best designed and developed organically and interactively with stakeholders within the context and setting in which it will be applied, and in alignment with the existing health, education, and technology enterprises.

According to Mintzberg and Lampel [7], the strategy literature began to unfold in the 1960s. Use of "strategy" development, once commonly applied by the private sector, has faded. Within the eHealth arena, high-level policy statements and "road-maps" are sometimes referred to as "strategy" but do not provide the evidence base and structure desirable for sustainable eHealth implementation. This current paper questions the value of eHealth for developing countries, demonstrates the need for eHealth strategy, and identifies three available tools, before enhancing one of these tools by embedding within it recognized strategy concepts and cognitive assessment approaches to create an enhanced eHealth Strategy Development Framework.

Growing expectations, changing demographics, and resource limitations require wise investment in eHealth solutions that address major health needs. Of even greater import, eHealth activities implemented now will establish the practice and technology infrastructure for decades to come. Sustainable eHealth solutions require development of a sound, evidence-based, and defensible eHealth strategy. Application of the enhanced eHealth strategy development tool presented here is recommended as a key initial step and presents health care institutions, subnational regions, or countries with a viable model.

eHealth Strategy in the Political and Policy Context

A desire exists to believe policy making is rational and based upon best available empirical evidence. Marmot [8] noted that the "evidence-based" movement attempted to influence the political/policy context to create more of an "evidence-based policy making" process, as opposed to making the evidence fit the political/policy context (termed "policy-based evidence making"). Within that frame, a very linear process was perceived: A policy issue would be identified, the scientist would gather the evidence, KT (knowledge transfer/translation) would ensure the evidence got to those who needed it, and decision-makers would inevitably make evidence-based decisions. However, examples suggest that policy making is not, in fact, based upon such a linear process or on the best available empirical evidence [9,10]; rather, it is often based on public opinion, electoral considerations, personal preference, and crisis management.

Some authors have taken a cautious, even cynical, view of evidence-based policy making and practice while others, however, point more optimistically to recent changes in attitude. For example, Fafard [11] states that just as evidence-based

medicine requires systematic analysis of available evidence, so too should evidence-based public policy be based on the careful testing of different policy and program options and notes that this is where the role of empirical evidence is the strongest. The author then concludes that two significant changes have occurred; first, there has been a shift from "evidence-based" to "evidence-informed" policy making, and second, there is renewed interest in taking into account the real life *context* of decision making.

It would seem that careful research is still required to make choices between an array of possible policy instruments and program interventions. This is particularly so in complex fields such as health, health care, and eHealth. The approach described in this paper ensures the evidence is provided and current context is thoroughly understood (including underlying values and value conflicts), and it therefore supports evidence-*informed* decision making regarding possible application of eHealth solutions.

Available Guidance for eHealth Strategy Development

Many developed countries (eg, Australia, EU countries) have established a variety of documents termed, or akin to, "eHealth strategy" [12,13]. They provide examples, but little or no guidance to the process of development. Furthermore, as described above, emulation or adaptation of approaches from elsewhere is not recommended. Recently Jones [14] published a strategy development guide that was specifically eHealth focused. While providing useful tools and guidance, it lacks theory and a holistic approach. In late 2012, the World Health Organization (WHO) and International Telecommunications Union (ITU) released their WHO-ITU National eHealth Strategy Toolkit [6], intended to provide a strategic framework and method for the development of a national eHealth vision, action plan, and monitoring and evaluation framework. This also provides useful tools and guidance, but its comprehensiveness may lead to complexity in its execution. Scott [15] first provided a framework for Strategic Planning in relation to eHealth, and it is that framework that provides systematic process, direction, and coherence, allowing any entity—regional, national, subnational, or facility—to develop its own eHealth strategy, leading to significant and measureable future impact.

Need for and Value of Developing an eHealth Specific Strategy

Is There a Need for eHealth Strategy Development?

eHealth in its largest sense has been practiced for many decades now—from basic telephony, through transmission of ECGs and images, to comprehensive e-records and even remote surgery. But despite this experience, there are few sustained eHealth implementations of demonstrated success and sustainability as evidenced through rigorous evaluation. The ITU [16] stated that, for at least the period 1960-2000, the "traditional cycle of telemedicine projects" was disappointing, and they noted that thousands of pilot sites, trials, tests, etc, took place but few of the initiatives survived beyond the end of their initial funding

period. They concluded that, during the 20th century, perhaps fewer than 10% of projects in developing countries were successful, with 45% faltering after just 1 year and the remaining 45% after 3 years. There is little reason to believe this has changed for initiatives implemented in the new century. Indeed, Ekeland et al [17] commented that available evidence on the value of telemedicine varies from "promising but incomplete" to "limited and inconsistent", with a particularly problematic area being economic analysis of telemedicine. Similarly, van Eland-de Kok et al [18] identified only small to moderate positive effects of eHealth on primary health outcomes of chronic disease patients and noted that due to the limited number of studies and methodological limitations, the evidence was not fully convincing.

A similar circumstance exists for large-scale electronic record initiatives, with large health informatics applications in developed countries failing to prove as successful as desired. For example, Electronic Health Records (EHRs) have been, or continue to be, introduced in many developed countries such as England, Scotland, France, Canada, Australia, and the USA—and at significant cost and risk. Originally budgeted at £2.3 billion, the United Kingdom is estimated to have spent between £6.2 billion [19] and £20 billion [20] on its NHS Connecting for Health program—and abandoned the program in 2011 as largely a failure [21]. Some estimates of Canada's pan-Canadian eHealth initiatives suggest a total expenditure of $10 billion [22] (with additional investment by provinces and territories), and questions of value have arisen in Ontario, British Columbia, and Alberta [23]. The bulk of these expenditures have been borne by the public sector, given that the private sector avoids investment until it sees a sound market opportunity. Black et al [24] completed a "systematic review of systematic reviews" of various eHealth solutions on the quality and safety of care and concluded that "despite support from policy makers, there was relatively little empirical evidence to substantiate many of the claims made in relation to these technologies". Also, Jamal et al [25] systematically reviewed the impact of health information technology (HIT) or health information systems (HISs) on the quality of health care and found insufficient evidence of either clinically or statistically important improvements in patient outcomes.

In regard to developing countries, Fernandez and Oveido [26] observed that, for the Caribbean region, it is only well-managed health institutions that plan medium- and long-term eHealth programs that are likely to be able to implement successful initiatives. According to these authors, ICT projects in the region are usually short-term and unsustainable, due to expectations of "instant results" and a lack of support for the new projects stemming from a lack of knowledge and understanding by policy and decision makers. They also highlight the lack of standardization needed to encourage the interjurisdictional sharing of information. These observations are likely to be equally applicable to institutions, as well as health systems, in most other developing countries and regions.

It would seem clear that our current approach to eHealth implementation does not work, and an alternate approach is needed.

Is eHealth a Viable Solution for Developing Countries?

The potential of eHealth to address growing health system concerns and health care needs is often identified in the literature, but clear evidence of its value remains uncertain. With these perspectives in mind, it is reasonable to ask "is eHealth a viable solution for developing countries?"

Despite the lack of success described above, there is some evidence from the developed world that HISs address health concerns and may lead to cost savings. But, even then, are the health concerns addressed by HISs in developed countries (eg, reduced adverse drug reactions) the most relevant to the developing world? Furthermore, the European Commission [27] found that for EHRs and ePrescribing in European countries, at least 4 years (more typically up to 9 years) are required to show positive annual socioeconomic return (SER), and 6-11 years to realize a cumulative net benefit. Given this time to realize SER in developed countries, can developing countries run the risk? Finally, are the projected cost savings for developed countries even feasible elsewhere? The United States spent an estimated US $8650 per capita (almost 18% GDP) on health in 2011, and Canada spent Can $5800 (projected) per capita (11.6% GDP) in 2011 [28]. In health systems that spend $6000 to $9000 per capita on health per year, perhaps there is room for savings through greater efficiencies. But in health systems that spend $10-35 per capita per year (as in many developing countries), are *any* cost savings likely? The business case is unlikely to be made through cost savings alone.

How Much Is Available to Spend on eHealth Solutions?

The WHO's Report of the Commission on Macroeconomics and Health [29] identified that countries needed to spend, at that time, a minimum of $34 per capita to provide just a basic health care package to their population. Introducing another element that requires funding, ie, eHealth, becomes an "opportunity cost". If you spend money on eHealth, you have to take it away from something else—immunization, sanitation, clean water, rural clinics, health provider salaries. Not all the funding will come from donors—sustainable solutions require investment by the country too. So how much does a developing country have available to spend on eHealth? To place this in perspective, consider the following. Of the Can $5800 per capita spent in 2011 by Canada, about 72% (OECD country average) or $4176 came from the public purse. Of this money, nearly 2.7% was spent on technology use in health (only some of which was eHealth), meaning Canada spends around Can $113 per capita on ICT use in health. In a country that spends $10-35 per capita on all its health needs, 2.7% would amount to 27-67 cents per capita on all technology applications. What eHealth solution can be bought, implemented, maintained, and sustained for that price?

What Is the Value of an eHealth Strategy?

So, is eHealth a solution for developing countries? Perhaps, but the solutions and approaches are unlikely to be those pursued in developed countries and must be aligned with the specific health system and health needs of the entity (institution, subnational region, country) and culture involved [30]. To achieve this very complex goal, an eHealth strategy is essential

to provide evidence-based guidance, describe the needs, and justify any expenditure, and thereby ensure wise investment of already incredibly scarce resources.

There are also synergistic effects. Once a national eHealth strategy is in place, it encourages (perhaps requires) facility-level eHealth strategy development, which aligns with and supports the national-level approach. Similarly, within a geographic or trading "region", countries can align their own approaches to develop a regional eHealth strategy. Several benefits are inherent in such an approach. Countries and regions take ownership of their own eHealth destiny and can guide (or decline) opportunities presented by external agencies. Furthermore, the shared experience allows more rapid accomplishment of sustainable eHealth implementations (see Figure 1).

Figure 1. Synergistic benefits derived from an eHealth strategy.
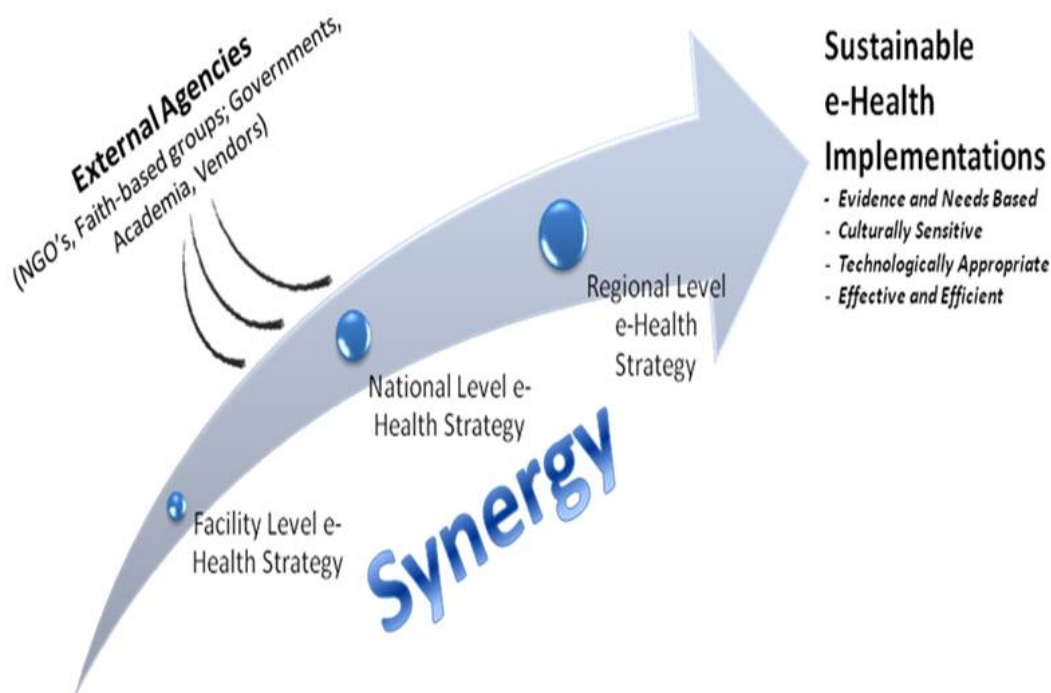
Invoking "Strategy" and "Complex System Analysis"

Strategy

Originally a military concept, strategy development became common in the private/business sector during the 1950s and 1960s. Since that time, overt strategy development has faded; some reasons cited include failure to differentiate and treat separately "direction-giving/leadership" and "managing", "irrational exuberance" of the markets, lack of respect for "direction-giving/leadership" as a profession, and failure to take seriously the need for "strategic thinking" and subsequent implementation in any learning organization [31]. Strategy development in the health sector (particularly for eHealth) is uncommon, yet much could be gained by recognizing the value of strategy development and its application to the eHealth environment.

Strategy in its simplest sense can be considered clarity around *where* you are going and *why* you are going there. According to Porter [32], strategy is creating fit among an organization's activities (without fit, there is no distinctive strategy and little sustainability), and the success of a strategy depends upon doing many things well and integrating them correctly. In the context of eHealth, an eHealth strategy would be documentation that describes the overall approach to be taken by an entity (institution, subnational region, country). It will identify and implement technologically appropriate and culturally sensitive eHealth solutions in the most appropriate manner and for the most appropriate purposes, explaining not just what is to be done, but why (given the prevailing circumstances). Strategy is key to sustainable eHealth implementation—indeed, the foundation for sustainability is strategy development. Many countries and organizations may claim to have an eHealth strategy (eg, the "Road Maps" of EU countries), but these tend either to be too narrow in focus or too general and abstract and often begin with a goal or an objective that is stated without

substantive context and perspective as to its rationale or origin, its impact on prevalent health needs, or any insight around its selection versus alternatives.

What approach to strategy development is most appropriate for the complexity and continuously developing eHealth setting? Boisot [33] has presented a typology that describes four different kinds of approach to strategy determined by the level of "turbulence" and "understandability" of the setting (Figure 2). According to Boisot, *intrapreneurship* is a state of great unpredictability and flux where entities respond as best they can under the chaotic circumstances surrounding them; *emergent strategy* is the product of "top down" and "bottom up" approaches which emerge incrementally over time without focussed effort; and *strategic intent* is an intuitively clear direction that can be pursued despite the turbulence present and

that permits activities to be aligned with a common purpose. Finally, *strategic planning* is viewed as formal consideration of a future course and has value in forcing consideration of two primary factors—the country's setting, and the inherent uncertainty surrounding eHealth. In this way, the strategic planning process matches appropriate activities to the evolving eHealth environment.

eHealth is recognized to be a constantly evolving field, but the turbulence that existed in the early days has passed. Similarly, sufficient research and application has taken place that sound lessons and good "understandability" exists of where and how to apply eHealth. Thus, within Boisot's typology, "strategic planning" lies at the intersection of high understandability and low environmental turbulence and is the appropriate strategic option to pursue.

Figure 2. Boisot's typology of strategy options.

	UNDERSTANDABILITY of the ENVIRONMENT	
	LOW	HIGH
HIGH LEVELS OF ENVIRONMENTAL TURBULENCE	Intra-preneurship	Strategic Intent
LOW LEVELS OF ENVIRONMENTAL TURBULENCE	Emergent strategy	Strategic Planning

Complex System Analysis

What makes a setting complex? Often it is the presence of a large number of interconnected parts whose interaction is not merely additive (a "simple" setting), but synergistic where the combinations and permutations are large and the outcomes not always obvious considering the properties of individual components. Complexity can be disorganized or organized [34]: disorganized complexity arises merely through the presence of a very large number of component and interconnected parts; organized complexity arises because the interconnected parts

exhibit emergent properties—complex patterns arising out of a multiplicity of relatively simple interactions between even a small number of parts. Human economies, social structures, health systems, and ICT infrastructures are all considered complex settings. It follows therefore that eHealth—with a large and growing number of potential applications (eg, technological and medical/health options), each of which interact with one another creating a complex setting—is certainly complex.

How can such complex settings be assessed? Traditionally, we strive to reduce complex systems to simpler subparts and analyze

those. However, in doing so, it can be argued that the "real-life" context and relevance of any analysis is lost. Other approaches are required. One approach is to examine settings in a holistic manner, which (according to the Oxford dictionary) is "characterized by comprehending the parts of something as intimately interconnected and explicable only by reference to the whole". Holistic analysis is typically interdisciplinary, concerned with the behavior of complex systems and respects occurrence of "feedback" (ie, when information about an event in the past will influence an occurrence (or occurrences) of that same or related event in the present or future).

Another approach is to mimic something we do innately (eg, when driving) to understand complex and dynamic settings, that is, create "situation awareness" (defined as "the perception of elements in the environment within a volume of time and space, the comprehension of their meaning, and the projection of their status in the near future" [35]). Situation awareness

arises when elements within the immediate setting are clearly perceived with respect to time and/or space, their meaning is comprehended, and projections are made of their status within the setting after some variable has changed (eg, time, speed, direction). It is an accepted tool for critical decision making in complex, dynamic areas [36], since current awareness determines what issue(s) are addressed next as well as interpretation of the information perceived [37]. The process by which this is done is termed "situation assessment" (sometimes "situational assessment") and is a form of tactical analysis that can be related to strategic and scientific analysis as seen in Figure 3.

Combining the approaches of holistic review and situational assessment, performance of a "holistic situation assessment" is recommended and is embedded within the enhanced eHealth strategy development framework described below.

Figure 3. Relationship among several cognitive processes.

Objective (Timeline)	Process	Output
Tactical (Immediate)	Situation Assessment	Situation Awareness
Strategic (Intermediate)	Sense Making	Understanding
Scientific (Long-Term)	Understand	Predict

Principles of eHealth Strategy Development

Before considering development of institutional, regional, or national eHealth strategy and policy, there are some fundamental principles that need be adopted. These are outlined below.

Principle 1: Simplify Complex Contexts

Experience gleaned from the literature shows the process of integrating eHealth as a routine health care tool faces many challenges, is very complex, and requires significant time. However, by establishing a sound and evidence-based eHealth strategy, it is possible to reduce the impact of such realities. The process is most effective when undertaken by a local (institutional, regional, country) team, as it builds local capacity, is designed by those most intimately knowledgeable about the

setting, and establishes pride and commitment of ownership for the undertaking and product.

Principle 2: A Pragmatic Approach Is Best

The goal of the strategy is to find an optimal solution to the most pressing (existing or anticipated) health-related problems. In other words; the approach is very focused, very health or health care "needs-based", and strongly "evidence-informed", but not overly researched (see Step 1 in the Process section below). This requires an understanding of pressing health care needs and alignment with, or creation of, a clear eHealth strategy to address them.

Principle 3: Spread the Cost

Networking provides opportunities to spread the cost of infrastructure and "infostructure" development between the government, business, agriculture, education, and health sectors.

For example, the ICT network supports all these sectors, not just one, and therefore the cost burden should not be borne by just one sector.

Principle 4: Balance Which eHealth Components Are Applied

Four primary components of eHealth exist [38]: (1) health informatics (collection, analysis, and distribution of health related data; eg, electronic records, surveillance), (2) telehealth (direct or indirect interaction with other health care providers, ill patients, or well citizens, eg, teleconsultation; social networking), (3) e-learning (use of ICT to provide teaching and education opportunities to health care providers and citizens), and (4) e-commerce (related to the business side of health care, eg, electronic reimbursement).

Solutions to specific health issues may require a predominance of one component over others, but it is likely any sustainable and comprehensive solution will require elements of each.

Principle 5: eHealth Solutions Must Be Right for the Setting

eHealth solutions that are identified for implementation should be technologically appropriate and culturally sensitive.

Appropriate technology can be defined as the most benign technological solution that achieves the desired purpose within the confines of current social, cultural, environmental, and economic conditions of the setting in which it is to be applied and that promotes self-sufficiency on the part of those using it in that setting. Described in this fashion, an appropriate technology would typically be simple to adopt and require fewer resources to operate and maintain (making it more likely to be sustainable and environmentally friendly).

Cultural sensitivity requires solutions to respect local traditions, expectations of the health care system, beliefs about health and disease, and patterns of usage of available health care services. Ignoring local health culture, such as traditional medicines or influential shamans, may undermine efforts to introduce eHealth initiatives. Or insufficient local resources may lead to abuse of modern medicines, such as using reduced doses of antibiotics, which may permit development of resistant strains capable of global spread. Solid experience and knowledge of cultural limitations must guide the design and implementation of eHealth solutions [39].

Principle 6: Provide Long-Term Focus

A clear, broadly accepted vision is required to guide the process, and garner sustained support from diverse stakeholders (eg, "eHealth facilitated health care by 2020"; "Integrated eHealth-care in 5 years").

Principle 7: Provide Medium-Term Targets

Enunciating a specific goal that people can embrace helps build and maintain momentum, for example, "To establish a needs-based, evidence informed, and national 5-year eHealth strategic plan that adopts technologically appropriate and culturally sensitive eHealth solutions and guides eHealth policy development".

Process of eHealth Strategy Development

Developed originally as a Telehealth Strategy Development Framework [15], this tool has been adapted for eHealth and further enhanced by embedding strategy and cognitive process theory and approaches (described above). Identification of specific methods and processes for collecting, managing, and using the information gathered during implementation of the tool continues. Assuming the above seven principles have been embraced and employed, there are seven steps to development of an eHealth strategy (the 8th step), which then guides and informs further undertakings, including the 9th step (policy development), and subsequent steps, eg, design of an enterprise architecture plan, business plan, readiness assessment plan, implementation and change management plan, operational plan, evaluation plan, and so forth. The seven steps are described below.

Evidence Gathering and Situation Assessment (Step 1)

To be effective, the eHealth strategy must address those specific health issues of most importance to the entity developing its eHealth strategy. Information regarding this will already be available in country/institution, NGO, or international agency reports (eg, WHO's annual country health status reports), local or regional planning documents, administrative databases, as well as through literature review. The available information can be interrogated to reveal insight regarding what the issues are; what the causes and/or contributing factors are to each issue; how serious (size, scope) each issue is; who is impacted by each issue and where they are located; how many are impacted by each issue; what community/population characteristics may be related to each issue; what has been done in the past to address each issue; and why the interventions succeeded or failed.

In this way, the reports/literature (the evidence) will have identified the specific health issues (the needs) that must be addressed and allowed some analysis of any linkages between sociodemographic features, and health indicators, health risks, and service use. The process may also have revealed information gaps that may require addressing. This evidence gathering and situation assessment step establishes a sound foundation and baseline that is defensible to critics and also provides a preliminary list of areas where an eHealth application *may* offer a solution.

Holistic Review (Step 2)

Overview

At this point, holistic situation assessment begins. It is necessary to examine many factors beyond just health needs to see if they guide decisions in a certain direction or identify potential barriers to some presumed solutions (this holistic approach has been used in other settings [40]). The goal is to examine the broader socioeconomic, political, and environmental context in relation to their impact on health need and to identify available assets, strengths, and capacity that might be brought to bear on the identified issues.

Although not an exhaustive list, information regarding each of the following examples will impact eHealth-related decisions.

Poverty (Spatial Distribution)

eHealth is considered by some as a tool for increasing equity of access to health care; mapping where poor and other vulnerable populations are located in relation to available health care facilities aids understanding of where implementation of eHealth solutions may be most beneficial [41].

Economic Policy Framework

ICT innovation continues to evolve, with new applications impacting all aspects of our economies and societies. Some reports suggest the public sector can begin large, expensive ICT projects without a clear understanding of goals, required resources, or risks. Understanding if and how government investment and policy formulation impacts ICT innovation, including eHealth, is essential [42].

Physical Geography

eHealth is recognized to remove or at least mitigate the barriers of time and space that physical geography imposes; however, technological solutions may also be limited by geography. Thus, solutions suitable for open expanses (eg, coastal areas, deserts) may not be suitable for extremely mountainous areas, highlighting the potential need for investment in different communication and technological eHealth solutions in different regions [43]. Consequently issues such as "line of sight" solutions, practical limits to wireless connectivity without repeater sites, location of infrastructure in disaster prone areas, etc, may restrict options.

Governance Issues and Policy Stability

Available experience with e-government, and the strength of the local supportive setting for eHealth will influence the acceptability and implementation of eHealth solutions. Furthermore, long-term vision, planning, and continuity in implementation despite political change is critical to success [30].

Cultural Barriers

Culture influences health care in several ways, including preferences for different treatments, individual health beliefs, and attitudes toward disclosure of medical information; eHealth solutions must be culturally appropriate if they are to be adopted and sustained [44,30].

Geopolitics

Factors such as geography, economics, and demography influence the politics, especially the foreign policy, of a country, which can influence intra- and interjurisdictional eHealth (eg, could neighboring countries support/share eHealth infrastructure or initiatives?)

Resource Issues (Including Human Health Resources)

Availability and skill set of the current cadre of health care (and eHealth) providers must be built to a critical point if countries are to "build the capacity to build their own capacity" [45].

e-Readiness or eHealth Readiness

Readiness to succeed in adoption, implementation, and use of any technology solutions is critical. The same is so for eHealth and the level of readiness of the public, health care providers, and the government must be thoroughly assessed (reassessed) to reveal gaps requiring intervention [46].

Linkages

NGOs or other agencies have become indispensable in the delivery of health in many developing countries [47]. However, these entities (eg, International Development Research Centre (IDRC), Swedish International Development Cooperation Agency (SIDA), Canadian International Development Agency (CIDA), International Red Cross, United States Agency for International Development (USAID), and faith-based organizations, etc) each have their own mandates, including perhaps eHealth strategies and activities. Without coordination and linkage of their activities to a specific eHealth strategy, their activities are, at best, ad hoc and confusing, and at worst, counterproductive, even detrimental.

Infrastructure

Fulmer [48] described infrastructure as the physical components of interrelated systems providing commodities and services essential to enable, sustain, or enhance societal living conditions. Given that eHealth is an ICT-based solution, the availability, type, capacity, coverage, cost, and location of current and planned physical ICT infrastructure will significantly influence the type and sophistication of eHealth solution feasible. Conversely, an eHealth Strategy may also inform discussion around just what type, capacity, coverage, cost, and location of planned infrastructure is needed.

Infostructure

In contrast to the "physical components", infostructure can be described as those human resources, organizational and administrative structures, policies, regulations, and incentives that facilitate fully integrated and sustainable use of innovative ICTs and services to improve health care in an organized response to health and health care needs, issues, and challenges (ie, eHealth). Once again, completing an eHealth Strategy will also inform discussion around each of these issues.

The holistic review must also consider distant or unpredictable events (eg, climate change, humanitarian disasters, natural disasters), so that implemented solutions have sustainability and flexibility. Climate change remains debated, but whether natural or iatrogenic, some locations are predicted to turn from lakes into deserts or erase low-lying islands or traditional residential regions in estuaries, which will impact population movement and perhaps distribution of diseases. Humanitarian disasters (drought, war) can cause mass migration of populations across borders, stretching still further already stretched health care systems. Finally, natural disasters such as floods or earthquakes can cause extensive damage or destroy ICT infrastructure, and such considerations should impact the type and location of ICT infrastructure adopted during the strategy development process.

Individually and collectively, each of these factors has a bearing on the type of solution (and eHealth solution) that might be most appropriate and most sustainable for any identified health need and population. "Mind mapping" software (which creates diagrams of relationships between concepts, ideas or other pieces of information) is a valuable tool to assist in this process.

This holistic situation assessment step is crucial but is not typically or overtly undertaken. Absence of such a sound review undermines the credibility of any subsequent eHealth strategy, questions fiscal responsibility, and will adversely affect the sustainability of proposed applications.

Differential Diagnosis (Step 3)

This is a tool taught to physicians during their training: "a systematic method of diagnosing a disorder that explains presenting signs and symptoms of a patient". But what happens if 2 patients appear with similar signs and symptoms? Do they have the same disease? Maybe, but not necessarily. The signs and symptoms of 2 or more patients may be similar, but the actual diagnosis can be different and differential diagnosis allows this to be resolved. Because the diagnoses for the 2 patients are different, so too will be the *treatment* and management for each patient. This is analogous to assessing the health needs of different institutions, subnational regions within a country, or countries. The health *issues* and *settings* may be similar, but when examined carefully (holistic situation assessment), the real health needs (and possible eHealth solutions) are seen to be different. Sachs [40] applied this differential diagnosis approach to his economic assessment, and the same principle is applied here to differentiate possible solutions.

Using the data, information, and analysis garnered in Steps 1 and 2, it is possible to look at groupings at the next level down (eg, subnational entities for a national eHealth strategy, districts/wards for a regional eHealth strategy, and communities for an institutional eHealth strategy) and to reveal differing needs of distinct locales or populations. These should be highlighted for later consideration.

Preliminary Prioritization (Step 4)

Given resource limitations, not every option can be pursued; trade-offs are essential and enforcing choice purposefully limits the options. But how do you choose? Priorities in health needs are traditionally viewed only in terms of disease morbidity and mortality. While intended for setting research priorities, the explicit and rational approach of the Combined Approach Matrix [49,50] takes into account other relevant determinants and can be used to prioritize the identified health needs. It consists of 5 different sources of evidence to formulate a priority list. These sources are (1) disease burden, (2) determinants, (3) level of knowledge, (4) economic cost, and (5) resources. Alternative, and perhaps more objective, tools are available (Sum of Ranking Approach (SRA), and Product of Value Approach (PVA)) as applied by the Ministry of Health and Long-Term Care (MOHLTC) in Ontario, Canada [51]. However, availability of sufficient quantity and quality of data for each health need may be a challenge. The overall goal of this step is to determine priority health needs and their associated characteristics for further review.

Identifying Solutions (Step 5)

At this point in the process, the evidence-informed, and needs-based, health issues of the institution, region, or country are known and have been prioritized. In addition, the internal and external influences that the current and future setting may bring to bear are also understood. It is now possible to consider a variety of solutions to address these identified health issues. But it is essential that expansive thought be employed. These solutions need *not* involve technological intervention and might function at one or more of the practice, process, or policy levels.

This stage is the point at which to engage a broad selection of local (institution, region, or country) stakeholders, including government, private sector, and academic participants with diverse experience and expertise in health, education, and business to become an eHealth strategy advisory team. The group must be briefed using the material gathered and analyzed in Steps 1-4, thereby creating a well-informed and up-to-date team. Their task is to assess the identified and prioritized health needs, consider the political context, leverage existing (or recommend potential) partnerships, and develop innovative solutions to the top 20% of the prioritized health needs (note that innovation is often considered synonymous with the use of sophisticated technological solutions—this is *not* the case; a dictionary definition of innovation is simply something that is "new to you"). A secondary, but crucial, purpose of establishing this team is to begin the process of intrajurisdictional capacity building and developing a knowledgeable eHealth strategy culture.

Considering eHealth Solutions (Step 6)

Only at this stage is the possible application of eHealth interventions considered. This process is best undertaken with the assistance of local or (if insufficient in number or expertise) external eHealth experts (telehealth, health informatics, e-learning, and e-commerce), who then become a part of the local working group and are briefed on both the prior material (Steps 1-4) and the process and solutions identified in Step 5. Again, expansive thinking is essential; many eHealth solutions are available but each may be optimal for only specific settings. It is recommended that attention still be focused on the top 20%; eHealth solutions may well be feasible for the remaining 80%, but if they are not highly prioritized then funding such initiatives may not be the wisest investment.

Options must be limited to a small number, and for each proposed eHealth application a brief but structured review (essentially a summary "business case") must be prepared. This will help to assess the feasibility of each solution for the given institution, region, or country; not all of the proposed eHealth applications will be technologically appropriate, culturally sensitive, or financially feasible.

Secondary Prioritization (Step 7)

Almost invariably more than one eHealth solution is available for any specific need (eg, applying different technologies such as videoconferencing versus podcasts for CME of clinicians),

and more than one need can be addressed using eHealth (eg, telehealth consultation services to remote communities versus introduction of a public health surveillance tool). Decisions must be made. The business case analysis will have provided insight regarding potential cost, complexity of implementation, likely readiness to implement, and proportion of the population impacted—these features can be used to rank options as described earlier. In the absence of sufficient data to permit objective prioritization, then a more subjective approach will have to be taken. Here, each potential solution can be classified into applications that are considered *essential* to have, versus those that would be *good* to have, versus those that might be *nice* to have. eHealth solutions that address a high priority health need of modest or low cost and complexity, and impact a large proportion of the population would be optimal and identified as essential. This is a critical stage in the eHealth strategy development process, as it sets direction for allocation of resources and commits to a certain path of ICT infrastructure development and policy need.

Strategy Formulation (Step 8)

To create the institution, region, or country "eHealth strategy", the findings from Steps 1-7 are synthesized, and the recommended priority needs and selected eHealth solutions described. This eHealth strategy document then informs further action. It will guide the building of the necessary enterprise architecture, ICT infrastructure, processes, and policy environment, as well as the subsequent design, readiness assessment, implementation plan, change management plan, evaluation study, and sustainability program for the selected eHealth applications.

Policy Development

There is continued debate about strategy versus policy. While there may be no clear cut answer to "what comes first—the chicken or the egg?", logically it follows that poor inputs to designing an eHealth favorable policy environment will ultimately result in poor outputs, poor outcomes, and undesirable impacts. Consequently, this paper is intended to encourage development of a sound and evidence-based eHealth strategy for any entity (region, country, subnational jurisdiction, or health care facility) as the *first* step: the premise being that strategy defines *where* and *why* action should be taken, whereas policy describes and implements *how* that action should be taken. During the strategy development process described above, barriers and facilitators to implementing the planned eHealth strategy will have been identified and documented highlighting specific areas of policy need in an evidence-based fashion.

Approaching eHealth policy development in this way ensures that important issues requiring policy solutions (new, revised, or rescinded policy) are identified. Attention can then be given to considering what specific eHealth policy is required to encourage and or manage the strategy, including expected growth in implementation and evaluation of eHealth solutions and sustainability. Such policy must be developed through an iterative, collaborative, and participatory process if support is to be engendered from all stakeholders. Further, it must be remembered that eHealth specific policy is only developed where it is not possible to achieve the desired result through

revision/amendment of existing health, education, or ICT policy. In the end, eHealth should become just another tool by which to provide health-related information, education, and services—to do so it must become integral to the existing health care system, not separate from it.

Discussion

eHealth Strategy Development

In our culture of constant growth, more and more governments and decision makers (eg, senior managers of health care facilities) are being called to task to demonstrate the value of the decisions they make. Health seemingly consumes a greater and greater proportion of available funds in an attempt to address the complex health care issues that plague all countries as we continue the global transition from infectious to noninfectious and chronic disease and old age. eHealth—a relatively new approach available to decision and policy makers in the developing world—has been hailed by many as a solution to these woes. Yet attempts to date in the developed world have shown relatively little success or return on investment despite significant outlay. A better and more reasoned means of understanding where and how to apply eHealth solutions is necessary. An evidence- and needs-based, transparent, and defensible eHealth strategy is required by each region, country, and facility.

Only recently has some guidance for eHealth strategy development become available. In 2011, the Commonwealth provided many templates and a structure to use in initiating a series of workshops intended to lead towards development of an eHealth strategy [14]. In that document, a good job was done of encouraging broad understanding of eHealth options as potential solutions (eg, Table 11 in [14]) and projecting future costs (eg, Table 12 in [14]). However, the toolkit does not describe processes for holistic situation assessment nor for prioritization of both health needs and eHealth solutions. More recently, the WHO/ITU provided a comprehensive document (a National eHealth Strategy Toolkit) [6]. Although promoted as a tool by which to create eHealth strategy, the content does not deliver a strategy per se, but rather provides guidance to achieve three outputs: (1) a National eHealth Vision, (2) a National eHealth Action Plan, and (3) Monitoring and Evaluation processes. Indeed, the document does not clearly distinguish between vision and strategy, nor seemingly identify development of an eHealth strategy as a specific output. It is certainly a comprehensive and valuable document, with much guidance provided on many steps to be taken after eHealth strategy development is complete.

While both these documents contribute significantly to the debate, neither provides conceptual background or theoretical support to the need for, and development of, eHealth strategy, nor do they focus on development of an eHealth strategy as a distinct and primary undertaking. Instead, both intermingle many other issues (eg, interoperability, standards, confidentiality, security, policy, architecture, implementation, change management, investment, benefits realization), which serves only to distract from the primary intended goal of eHealth strategy development. All of these aforementioned topics are

XSL•FO
RenderX

certainly of relevance but should be addressed only once a clear understanding has been developed of "where you want to go" and "why you want to go there", that is, having established the eHealth strategy.

Perhaps the greatest contribution of the eHealth Strategy Development Framework is its clear focus on establishing an evidence-based eHealth strategy and providing the conceptual understanding and tools required by which to achieve that. To be effective, an eHealth strategy must be solidly grounded in an understanding of the broader context within the setting (region, country, facility), and the challenges and opportunities that exist. It must provide clarity around the health need(s) that must be addressed and the solutions (particularly eHealth solutions) it is intended to apply.

The eHealth Strategy should not be so detailed and unwieldy that it cannot be used as a functional and guiding document. Therefore, it does not provide details of specific undertakings; those needs are addressed through next steps, including design of an enterprise architecture plan, business plan, readiness assessment plan, implementation plan, change management plan, evaluation plan, and operational plan. Once established, the eHealth Strategy acts as a pole star, that is, a "constant" to which all can refer as they work to achieve identified goals and navigate the defined path.

Conclusion

Growing expectations, changing demographics and disease patterns, and resource limitations require wise investment in eHealth solutions that address major health needs in any given setting. Solutions that are designed and implemented now must form the foundation (practice and technology infrastructure) for decades to come. Such sustainable eHealth solutions first require development of a sound, evidence-based, transparent, and defensible eHealth strategy, which then informs subsequent development of a sound and viable policy environment, enterprise architecture, and so forth. This paper describes the conceptual understanding and practical steps required for any facility, country, or region to develop its own eHealth strategy.

Authors' Contributions

RES conceptualized the framework. Both authors were responsible for the literature review and analysis. Both authors were involved in writing and reviewing the manuscript.

Conflicts of Interest

None declared.

References

1. World Health Organization. Health Topics - e-Health.: WHO; 2012 Nov. URL: http://www.who.int/topics/ehealth/en/ [accessed 2013-07-23] [WebCite Cache ID 6BsSSwlDP]
2. Scott RE, Mars M, Hebert M. How Global is e-Health and Knowledge Translation? In: Ho K, Jarvis-Selinger S, Novak Lauscher H, Cordeiro J, Scott RE, editors. Technology Enabled Knowledge Translation for eHealth: Principles and Practice (Health care Delivery in the Information Age). London: Springer; 2012:339-357.
3. Pan American Health Organization. eHealth Strategy and Plan of Action (2012-2017) - CD51-13. Washington, DC; 2011. URL: http://new.paho.org/ict4health/index. php?option=com_content&view=article&id=54%3Aestrategia-y-plan-de-accion-sobre-esalud-2012-2017&catid=15%3Aops-wdc&lang=en [accessed 2013-07-23]
4. Ministry of Medical Services, Ministry of Public Health and Sanitation. Kenya's National e-Health Strategy 2011-2017.: Government of Kenya; 2012 Jun. URL: http://www.medical.go.ke/images/stories/gallery/documents-2010-2011/kenyanation_ehealth_strategy.pdf [accessed 2013-07-22] [WebCite Cache ID 6IKk5wGQ4]
5. South Africa Department of Health. e-Health Strategy South Africa 2012-2016. 2012 Jul. URL: http://www.hst.org.za/sites/default/files/eHealth_Strategy_South_Africa_2012-2016.pdf [accessed 2013-07-22] [WebCite Cache ID 6IKkD4tRl]
6. World Health Organization, International Telecommunication Union. National e-Health Strategy Toolkit. Geneva; 2012. URL: http://www.itu.int/ITU-D/cyb/events/2012/e-health/Nat_eH_Dev/Session%201/ITU-WHO%20National%20eHealth%20Strategy%20toolkit%20March%202012.pdf [accessed 2013-07-22] [WebCite Cache ID 6IKkOLlHx]
7. Mintzberg H, Lampel J. Reflecting on the strategy process. Sloan Management Review 1999;40(3):21-30.
8. Marmot MG. Evidence based policy or policy based evidence? BMJ 2004 Apr 17;328(7445):906-907 [FREE Full text] [doi: 10.1136/bmj.328.7445.906] [Medline: 15087324]
9. Dopson S, Fitzgerald L. Knowledge to Action: Evidence-based Healthcare in Context. New York: Oxford University Press; 2005.
10. Nutley SM, Walter I, Davies HTO. Using Evidence: How Research Can Inform Public Services. Bristol: Policy Press; 2007.
11. Fafard P. Evidence and healthy public policy: Insights from health and political sciences. Canadian Policy Research Networks 2008.

XSL•FO

RenderX

12. National e-Health and Information Principal Committee. National eHealth Strategy. 2008. URL: http://www.health.gov.au/internet/main/publishing.nsf/Content/604CF066BE48789DCA25751D000C15C7/$File/Summary%20National%20E-Health%20Strategy%20final.pdf [accessed 2013-07-23] [WebCite Cache ID 6IKksmnRK]

13. Stroetmann KA, Artmann J, Stroetmann VN. Country Reports Database. 2012. URL: http://ehealth-strategies.eu/database/database.html [accessed 2013-07-24] [WebCite Cache ID 68oR2uWO9]

14. Jones T. Developing an e-Health Strategy: A Commonwealth Workbook of Methodologies, Content and Models. London: Commonwealth Secretariat; 2011.

15. Scott RE. Future proofing telehealth in developing countries. J Telemed Telecare 2007 Dec 01;13(8):70-72. [doi: 10.1258/135763307783247437]

16. International Telecommunication Union (ITU). Implementing e-Health in developing countries: Guidance and principles. Geneva: ITU; 2008. URL: http://www.itu.int/ITU-D/cyb/app/docs/e-Health_prefinal_15092008.PDF [accessed 2013-07-23] [WebCite Cache ID 6IKlOHj1O]

17. Ekeland AG, Bowes A, Flottorp S. Effectiveness of telemedicine: a systematic review of reviews. Int J Med Inform 2010 Nov;79(11):736-771. [doi: 10.1016/j.ijmedinf.2010.08.006] [Medline: 20884286]

18. Eland-de Kok P, van Os-Medendorp H, Vergouwe-Meijer A, Bruijnzeel-Koomen C, Ros W. A systematic review of the effects of e-health on chronically ill patients. J Clin Nurs 2011 Nov;20(21-22):2997-3010. [doi: 10.1111/j.1365-2702.2011.03743.x] [Medline: 21707807]

19. National Audit Office (NAO). The National Programme for IT in the NHS. 2006. URL: http://www.nao.org.uk/report/department-of-health-the-national-programme-for-it-in-the-nhs/ [accessed 2013-07-23] [WebCite Cache ID 6IKlWxBPr]

20. Fleming N. The Telegraph. 2012. Bill for hi-tech NHS soars to £20 billion URL: http://www.telegraph.co.uk/news/uknews/1473927/Bill-for-hi-tech-NHS-soars-to-20-billion.html [accessed 2013-07-24] [WebCite Cache ID 68oRIstp3]

21. Brewster T. IT Pro. 2012. NHS £11bn IT project officially ditched URL: http://www.itpro.co.uk/636292/nhs-11bn-it-project-officially-ditched [accessed 2013-07-24] [WebCite Cache ID 68oRYPIcP]

22. Auditor General of Canada. Electronic Health Records in Canada - An Overview of Federal and Provincial Audit Reports. 2010. URL: http://www.oag-bvg.gc.ca/internet/english/parl_oag_201004_07_e_33720.html [accessed 2013-07-23] [WebCite Cache ID 6IKlqvQDB]

23. Dal Molin J. Open Innovation: Transforming Health Systems through Open and Evidence Based Health ICT Innovation. Communications and Strategies 2011;83:17-36.

24. Black AD, Car J, Pagliari C, Anandan C, Cresswell K, Bokun T, et al. The impact of eHealth on the quality and safety of health care: a systematic overview. PLoS Med 2011;8(1).

25. Jamal A, McKenzie K, Clark M. The impact of health information technology on the quality of medical and health care: a systematic review. HIM J 2009;38(3):26-37. [Medline: 19875852]

26. Fernandez A. In: Oviedo E, editor. eHealth in Latin America and the Caribbean: Progress and challenges. Santiago, Chile: United Nations; 2011.

27. European Commission. The socio-economic impact of interoperable electronic health record (EHR) and ePrescribing systems in Europe and beyond. Final Study Report. 2009. URL: http://www.ehr-impact.eu/downloads/documents/EHRI_final_report_2009.pdf [accessed 2013-07-23] [WebCite Cache ID 6IKmTeJ0i]

28. Canadian Institute for Health Information (CIHI). Health spending to reach $200 billion in 2011. 2012 Jun. URL: http://www.cihi.ca/cihi-ext-portal/internet/en/document/spending+and+health+workforce/spending/release_03nov11 [accessed 2013-07-24] [WebCite Cache ID 68oSkzaUl]

29. World Health Organization. Macroeconomics and health: investing in health for economic development. Report of the Commission on Macroeconomics and Health. Geneva: World Health Organization; 2001.

30. Hernández-Torre M, Montiel-Amoroso G, Pérez-Jiménez M, Dávila-Montemayor M, Voisinee C. Health Projects in Mexico: The Contribution of Tecnologico de Monterrey. In: Ho K, Jarvis-Selinger S, Novak Lauscher H, Cordeiro J, Scott RE, editors. Technology Enabled Knowledge Translation for eHealth: Principles and Practice (Healthcare Delivery in the Information Age). London: Springer; 2012.

31. Garratt B. Preface to the Second Edition. In: Developing Strategic Thought: A Collection of the Best Thinking on Business Strategy. London: Profile Books; 2003:xi-xv.

32. Porter ME. What is strategy? Harvard Business Review 1996 Nov 01 [FREE Full text]

33. Boisot M. Preparing for turbulence: The changing relationship between strategy and management development in the learning organisation. In: Garratt B, editor. Developing Strategic Thought: A Collection of the Best Thinking on Business Strategy. London: Profile Books; 2003:34-55.

34. Weaver W. Science and complexity. Am Sci 1948 Oct;36(4):536-544. [Medline: 18882675]

35. Endsley MR. Toward a Theory of Situation Awareness in Dynamic Systems. Hum Factors 1995 Mar 01;37(1):32-64. [doi: 10.1518/001872095779049543]

36. Endsley MR. Measurement of Situation Awareness in Dynamic Systems. Hum Factors 1995 Mar 01;37(1):65-84. [doi: 10.1518/001872095779049499]

37. Endsley MR, Garland DJ. Situation awareness analysis and measurement. Mahwah, NJ: Lawrence Erlbaum Associates; 2000.

38.   Scott RE. Globale-Health Policy - From Concept to Strategy. In: Wootton R, Patel N, Scott RE, Ho K, editors. Telehealth in the Developing World. London: Royal Society of Medicine Press; Feb 2009:55.

39.   Scott RE, Palacios MF. e-Health - Challenges of Going Global. In: Scott CM, Thurston WE, editors. Collaboration in Context. Calgary: Institute for Gender Research & Health Promotion Research Group, University of Calgary Press; 2003:45.

40.   Sachs J. The End of Poverty: Economic Possibilities for Our Time. London: Penguin Press; 2005.

41.   Gauci A. Spatial maps: Targeting & mapping poverty. London: United Nations. Economic Commission for Africa; 2005.

42.   Victorian Auditor-General's Office. Investing Smarter in Public Sector ICT: Turning Principles into Practice. 2008 Jul 30. URL: http://www.egov.vic.gov.au/victorian-government-resources/reports-victoria/investing-smarter-in-public-sector-ict-turning-principles-into-practice.html [accessed 2013-07-23] [WebCite Cache ID 6IKns3OIy]

43.   Santi P. Topology Control in Wireless Ad Hoc and Sensor Networks. Toronto: Wiley; 2005.

44.   Berger JT. Culture and ethnicity in clinical care. Arch Intern Med 1998;158:2085-2090. [doi: 10.1001/archinte.158.19.2085]

45.   Mars M. Building the capacity to build capacity in e-health in sub-Saharan Africa: the KwaZulu-Natal experience. Telemed J E Health 2012;18(1):32-37 [FREE Full text] [doi: 10.1089/tmj.2011.0146] [Medline: 22150714]

46.   Khoja S, Scott RE, Casebeer AL, Mohsin M, Ishaq AF, Gilani S. e-Health readiness assessment tools for healthcare institutions in developing countries. Telemed J E Health 2007 Aug;13(4):425-431. [doi: 10.1089/tmj.2006.0064] [Medline: 17848110]

47.   Sabri B, Siddiqi S, Ahmed AM, Kakar FK, Perrot J. Towards sustainable delivery of health services in Afghanistan: options for the future. Bull World Health Organ 2007 Sep;85(9):712-718 [FREE Full text] [Medline: 18026628]

48.   Fulmer J. PEI Infrastructure Investor. 2009. What in the world is infrastructure? URL: http://www.corridortrust.com/uploads/Infrastructure_Investor.pdf [accessed 2013-07-23] [WebCite Cache ID 6IKoHB25J]

49.   Ghaffar A, de Francisco A, Matlin S. The Combined Approach Matrix: A priority-setting tool for health research. Geneva: Global Forum for Health Research; 2012.

50.   Ghaffar Olifson A, Collins T, Matlin SA, Olifson S. The 3D Combined Approach Matrix: An improved tool for setting priorities in research for health. Geneva: Global Forum for Health Research; 2012.

51.   Jain M. Evidence-Based Planning: The Chronic Disease Priority Setting Decision Toolkit, Version 2.: Health System Strategy Division; 2008 Oct. URL: http://www.ontla.on.ca/library/repository/mon/23004/291210.pdf [accessed 2013-07-22] [WebCite Cache ID 68odYc0XK]

Abbreviations

CIDA: Canadian International Development Agency

CME: continuing medical education

ECG: electrocardiograph

EU: European Union

GDP: gross domestic product

HIS: health information system

HIT: health information technology

ICT: information and communications technologies

IDRC: International Development Research Centre

IRC: International Red Cross

ITU: International Telecommunications Union

KT: knowledge transfer/translation

MOHLTC: Ministry of Health and Long-Term Care

NGO: non-governmental organization

NHS: National Health Service

OECD: Organization for Economic Co-operation and Development

PAHO: Pan American Health Organization

PVA: product of value approach

SER: socioeconomic return

SIDA: Swedish International Development Cooperation Agency

SRA: sum of ranking approach

USAID: United States Agency for International Development

WHO: World Health Organization

Original Paper

An 8-Week Web-Based Weight Loss Challenge With Celebrity Endorsement and Enhanced Social Support: Observational Study

Melinda J Hutchesson[1], BND, PhD; Clare E Collins[1], BSc, Dip Nutr&Diet, Dip Clin Epi, PhD; Philip J Morgan[2], BEd(HPE) Hons, PhD; Robin Callister[3], B Pharm, MSc, PhD

[1]School of Health Sciences, Faculty of Health, Priority Research Centre in Physical Activity and Nutrition, Callaghan, Australia

[2]School of Education, Faculty of Education and Arts, Priority Research Centre in Physical Activity and Nutrition, University of Newcastle, Callaghan, Australia

[3]School of Biomedical Sciences and Pharmacy, Faculty of Health, Priority Research Centre in Physical Activity and Nutrition, University of Newcastle, Callaghan, Australia

Corresponding Author:
Melinda J Hutchesson, BND, PhD
School of Health Sciences
Faculty of Health
Priority Research Centre in Physical Activity and Nutrition
Hunter Building (HA12), University Drive
Callaghan, 2308
Australia
Phone: 61 49215405
Fax: 61 49217053
Email: melinda.hutchesson@newcastle.edu.au

Abstract

Background: Initial engagement and weight loss within Web-based weight loss programs may predict long-term success. The integration of persuasive Web-based features may boost engagement and therefore weight loss.

Objective: To determine whether an 8-week challenge within a commercial Web-based weight loss program influenced weight loss, website use, and attrition in the short term, when compared to the standard program.

Methods: De-identified data for participants (mean age 36.7±10.3 years; 86% female) who enrolled in the Biggest Loser Club (BLC) (n=952) and the BLC's Shannan Ponton Fast Track Challenge (SC) for 8 weeks (n=381) were compared. The BLC program used standard evidence-based website features, with individualized calorie and exercise targets to facilitate a weight loss of 0.5-1 kg per week (–500kcal/day less than estimated energy expenditure). SC used the same website features but in addition promoted greater initial weight loss using a 1200 kcal/day energy intake target and physical activity energy expenditure of 600 kcal/day. SC used persuasive features to facilitate greater user engagement, including offering additional opportunities for social support (eg, webinar meetings with a celebrity personal trainer and social networking) endorsed by a celebrity personal trainer. Self-reported weekly weight records were used to determine weight change after 8 weeks. A primary analysis was undertaken using a generalized linear mixed model (GLMM) with all available weight records for all participants included. Dropout (participants who cancelled their subscription) and nonusage (participants who stopped using the Web-based features) attrition rates at 8 weeks were calculated. The number of participants who accessed each website feature and the total number of days each feature was used were calculated. The difference between attrition rates and website use for the two programs were tested using chi-square and Wilcoxon Rank Sum tests, respectively.

Results: Using GLMM, including weight data for all participants, there was significantly greater (P=.03) 8-week weight loss in SC (–5.1 kg [–5.5 to –4.6 kg] or –6.0%) compared to BLC participants (–4.5 kg [–4.8, –4.2] or –5.0%). Dropout rates were low and consistent across groups (BLC: 17 (1.8%) vs SC: 2 (0.5%), P=.08) and 48.7% (456/936) of BLC and 51.2% (184/379) of SC participants accessed the website at 8 weeks, with no difference between programs (P=.48). SC participants accessed the discussion forums, menu plans, exercise plans, and educational materials significantly more than BLC participants (P<.05).

Conclusions: Using a short-term challenge with persuasive features, including online social support with endorsement by a celebrity personal trainer, as well as a greater energy balance deficit, within a commercial Web-based weight loss program may facilitate greater initial weight loss and engagement with some program components. The results support the need for a more

XSL•FO
RenderX

rigorous and prospective evaluation of Web-based weight loss programs that incorporate additional strategies to enhance initial weight loss and engagement, such as a short-term challenge.

Introduction

Recent systematic reviews suggest Web-based interventions facilitate modest weight loss [1-3], and participants' engagement with program features are a key factor associated with success [1]. Krukowski et al have shown that individuals who were consistent users of a Web-based weight loss program during the initial program weeks were more likely to continue to use the program features and achieve significantly greater weight loss after 6 months [4]. Furthermore, greater weight loss at the beginning of treatment has been identified as a predictor of long-term weight loss success and weight loss maintenance [5]. Therefore, Web-based weight loss programs that engage participants and enhance weight loss during the initial stages of treatment may be more successful in the long term.

Web-based weight loss program providers are therefore exploring new ways to improve initial program success. This includes the use of persuasive technology [6], which may positively influence participant engagement [7]. However, a previous review of the use of persuasive features by six popular weight loss websites indicated that techniques to date may not be very persuasive [8] due to poor dialogue support, limited credibility support, and moderate primary task and social support. This suggests that greater focus on evaluating the effectiveness of persuasive features in Web-based weight loss interventions, and how they influence engagement and weight loss success, is required.

Therefore, the aim of this observational study was to determine whether an 8-week "challenge" version of a commercial Web-based weight loss program influenced weight loss, website use, and attrition in the short term, when compared to the standard commercial Web-based weight loss program. The 8-week challenge provided enhanced system credibility support through the use of a celebrity personal trainer to endorse the program and host additional opportunities for social support.

Methods

Participants and Study Design

Participants were adults aged 18-74 years with a body mass index (BMI) >18.5kg/m^2 who subscribed to the standard Web-based weight loss program for at least the minimum subscription length of 12 weeks from June 27, 2011, to October 24, 2011, or the 8-week "challenge" version of the program, which began October 24, 2011. The subscription must have been the participants' first for the commercial program, and those who did not pay for their subscription (eg, free promotional program trials) were excluded. The cost of a subscription in 2011 was AU$149 for SC. For BLC, the cost ranged from AU$19.95 per month if paid upfront for 12 months to AU$49.95 per month if paid monthly.

Intervention

The commercial Web-based weight loss program was managed by SP Health Co, Australia. The standard program was The Biggest Loser Club (BLC) [9]. The short-term efficacy of the standard program, which is underpinned by social cognitive theory, incorporates key components of effective behavioral weight loss interventions, and includes persuasive features (Table 1), has been previously demonstrated [10]. The challenge version of the program was the Shannan Ponton Fast Track 8-Week Challenge (SC). Key program components are summarized in Table 1 and in Multimedia Appendices 1-4. SC included all features of BLC. However, to facilitate greater participant engagement SC drew on persuasive system design by offering enhanced system credibility support through the use of a celebrity personal trainer to endorse the program and host additional opportunities for social support. The celebrity personal trainer was Shannan Ponton, who is a qualified personal trainer on a national television program, "The Biggest Loser Australia". The SC also used more "challenging" energy intake and expenditure targets (Table 1), with the goal of achieving greater initial weight loss.

Table 1. Description of the key components of the Biggest Loser Club (BLC) and Shannan Challenge (SC) programs linked to Persuasive Systems Design 4 categories.

Component	BLC	SC
Diet and exercise recommendations		
	Individualized calorie targets based on participants' estimated total energy expenditure at enrollment based on their reported height, weight, and activity level and their desire to lose weight (–500kcal/day less than estimated expenditure) or maintain their weight [Primary Task Support].	Calorie target is 1200 calories for all participants. The exercise plan is 6 days/wk with aim to burn 600 calories/day [Primary Task Support].
Self-monitoring		
	Food and physical activity diaries to monitor calorie targets and search engines to facilitate entry of food data [Primary Task Support].	As per BLC
	Monitoring of reported body weight, waist and hip girths; graphical display of changes in data and body (BMI) silhouette. Participants were encouraged to "weigh in" once/wk [Primary Task & Dialogue Support].	As per BLC plus weight loss leader board with a prize for member who achieves the greatest percentage weight loss each week [Primary Task, Dialogue & Social Support].
Feedback		
	Daily and weekly calculations of energy balance and meeting recommended nutrient and food group targets from online diary [Dialogue Support].	As per BLC
	Automated (computer-generated) weekly personalized feedback on their dietary intake and exercise based on their diary entries, as well as their use or lack of use of the standard website features, and the level of success of their weight loss [Dialogue Support].	As per BLC
Education materials		
	Weekly menu plan and grocery list [Primary Task Support].	As per BLC
	Weekly physical activity plan [Primary Task Support].	As per BLC plus choice of a Home or Gym exercise program including video demonstrations from a celebrity personal trainer "Shannan" [System Credibility Support].
	Weekly tutorials, fact sheets, and challenges, which participants are prompted to access via a weekly email [Primary Task Support].	As per BLC
Social support		
	Discussion forum [Social Support]	Exclusive discussion forum where only Challenge members can post comments [Social Support].
	Historical online meetings hosted by an accredited practicing dietitian could be viewed by BLC participants [Social Support].	Weekly video blog with personal trainer "Shannan" [Social & System Credibility Support].
		Weekly online meeting with personal trainer "Shannan" including a video and chat function where members post questions and personal trainer replies in writing in real-time [Social & System Credibility Support].
		Facebook page where the personal trainer "Shannan" posts motivating messages, questions, or challenges [Social & System Credibility Support].
Reminders/Prompts		
	Participants received weekly reminders to weigh in via email or SMS [Dialogue Support].	As per BLC

Data Collection and Measures

SP Health Co collected the data that were provided to the researchers in a de-identified form. Ethics approval for the study was obtained from the University of Newcastle Human Research Ethics Committee, NSW, Australia.

Participants' pretreatment demographic (sex, age, and ethnicity) and anthropometric characteristics (weight and height) were captured from an enrollment survey. Participants' self-reported height and weight were used to calculate BMI (weight in kilograms divided by height in meters squared), which was

XSL•FO
RenderX

categorized as healthy (BMI 18.5-24.9 kg/m^2), overweight (BMI 25-29.9 kg/m^2), or obese (BMI ≥30 kg/m^2).

Participants self-reported weights (in kilograms) recorded online weekly were used to determine weight change achieved after 8 weeks. Two types of attrition rates were calculated: dropout and nonusage [11]. Dropout attrition rates were calculated based on the number of participants who did not complete the program and therefore included participants who cancelled their subscription prior to completing 8 weeks. Nonusage attrition rates were calculated based on the number of participants who did not drop out but stopped using all Web-based features and did not resume use within the 8 weeks. To describe website use, we calculated the number of participants who accessed each website feature (overall access, food diary entries, exercise diary entries, forum views and posts, menu plan, exercise plan, weekly educational materials, and live webinars attended) and the total number of days each feature was used.

Statistical Analysis

Data analysis was undertaken using Stata 11.0 (StataCorp). Basic descriptive statistics were used to describe the baseline characteristics and website usage data. Differences between the two programs were tested using chi-square test for categorical data, t tests if normally distributed continuous data, or Wilcoxon Rank Sum test if non-normal continuous data. To determine the weight change achieved from enrollment to 8 weeks, generalized linear mixed models (GLMM) were utilized containing available self-reported weight records for all participants. Baseline age and sex were controlled for in the analyses as potential confounders. A secondary sensitivity analysis was conducted to determine the robustness of the results from the GLMM approach, by imputing missing 8-week weight data using the last observation carried forward (LOCF) method. A linear regression model was fitted with weight at 8 weeks as the outcome variable, group as the predictor variable, and enrollment weight, age, and sex as covariates.

Results

Pretreatment Characteristics

Overall the study included 1334 individuals (953 BLC, 381 SC). Participants were predominantly female (86%), with an average age of 36.7±10.3 years, and half were of Anglo-Saxon descent (Table 2). At enrollment, SC participants had a lower mean BMI (30.6 vs 33.0 kg/m^2, P<.001) than BLC participants.

Attrition Rates

In total, 19 participants (1.4%) dropped out during the 8-week period with no significant difference in dropout rates between the two programs, but a possible trend of higher dropout rates among BLC participants: BLC 17 (1.8%) vs SC 2 (0.5%), P=.08. Nonusage attrition rates were 49.7% (n=653/1315) at 8 weeks. There was no significant difference in nonusage attrition rates between programs (P=.47) as 51.2% of BLC (480/936) and 48.8% of SC (185/379) participants stopped using all website features during the 8 weeks and did not return to use.

Weight Loss

The primary analysis using GLMM resulted in a mean self-reported weight reduction of –4.6 kg (95% CI –4.9 to –4.4 kg) or –5.3% for all participants. SC participants self-reported significantly greater weight loss (–5.1 kg [–5.5 to –4.6 kg] or –6.0%) than BLC participants (–4.5kg [–4.8, –4.2] or –5.0%) after 8 weeks (P=.03) with small effect sizes (0.06 and 0.14 respectively).

The sensitivity analysis using LOCF gave a mean self-reported weight loss of –2.7 kg (–2.9 to –2.6 kg) or –3.0% after 8 weeks for all participants. SC participants self-reported significantly greater weight loss (–3.0 kg [–3.3 to –2.6kg] or –3.4%) than BLC participants (–2.6 kg [–2.9 to –2.4 kg] or –2.9%) after 8 weeks (P<.001), with small effect sizes (0.09 and 0.15 respectively). A significantly higher proportion of SC participants self-reported a weight loss of ≥5% after 8 weeks than the BLC participants (27.3% vs 22.6%, P=.02). See Table 3.

Website Use

Website use for SC and BLC participants is described in Table 4. SC participants accessed and posted to the discussion form, viewed the program plan, menu plans, exercise plans, and educational materials more than BLC participants (Table 4). BLC participants made a significantly greater number of food entries and used the online diary on significantly more days than SC participants (Table 4). There were no significant differences in the number of weekly weigh-ins recorded or the number of exercise entries in the online diary between the two programs.

Table 2. Baseline characteristics of participants who enrolled in the BLC or the SC programs.

		Total (n=1334)	BLC (n=953)	SC (n=381)	P value[a]
Gender, % (n)					
	Female	85.9 (1146)	85.2 (812)	87.7 (334)	.22
Age in yrs, mean (SD)		36.7 (10.3)	36.5 (10.7)	37.1 (9.1)	.35
Ethnicity, % (n)					
	Anglo-Saxon	51.0 (680)	49.5 (472)	54.6 (208)	.55
	Other	18.8 (251)	19.9 (189)	16.3 (62)	
	Did not wish to respond	30.2 (403)	30.6 (292)	29.1 (111)	
BMI (kg/m^2), mean (SD)		32.3 (7.0)	33.0 (7.1)	30.6 (6.5)	<.001
	Normal weight % (n)	12.4 (166)	10.3 (98)	17.9 (68)	<.001
	Overweight % (n)	31.8 (424)	29.1 (277)	38.6 (147)	
	Obese % (n)	55.8 (744)	60.7 (578)	43.6 (166)	

[a]Differences between programs tested using chi-square test for categorical data and t tests for continuous data.

Table 3. Weight change after 8 weeks for BLC and SC participants using GLMM and LOCF (all analyses controlled for baseline age and gender).

	Mean (95% CI)				Effect size, Cohen's d	P value for difference between groups
	Total (n=1334)	BLC (n=953)	SC (n=381)	Difference between groups		
Primary analysis, GLMM						
Absolute (kg)	-4.6 (-4.9, -4.4)	-4.5 (-4.8, -4.2)	-5.1 (-5.5, -4.6)	-0.6 (-1.2, -0.6)	0.06	.03
Percentage (%)	-5.3 (-5.4, -5.1)	-5.0 (-5.2. -4.7)	-6.0 (-6.4, -5.7)	-1.1 (-1.5, -0.7)	0.14	<.001
Sensitivity analysis, LOCF						
Absolute (kg)	-2.7 (-2.9, -2.6)	-2.6 (-2.9, -2.4)	-3.0 (-3.3, -2.6)	-0.3 (-0.5, -0.2)	0.09	.005
Percentage (%)	-3.0 (-3.2, -2.8)	-2.9 (-3.1, -2.6)	-3.4 (-3.7, -3.0)	-0.5 (-0.7, -0.4)	0.15	<.001
Percentage weight change category						
Weight gain, % (n)	3.9 (52)	4.7 (45)	1.8 (7)	N/A	N/A	.02
0% to <5%, % (n)	72.1 (962)	72.6 (692)	70.9 (270)			
5% to <10%, % (n)	20.8 (278)	20.0 (191)	22.8 (87)			
10% or more % (n)	3.2 (42)	2.6 (25)	4.5 (17)			

XSL•FO
RenderX

Table 4. Website use by BLC and SC participants from enrollment to 8 weeks.

	BLC (n=953)		SC (n=381)		P value	
	Percentage used, % (n)	Frequency[a] (median IQR)	Percentage used, % (n)	Frequency[a] (median IQR)	Percentage used[b] %	Frequency[c]
Weekly weigh-in	97.8 (932)	5 (2-8)	97.6 (372)	5 (2-8)	.86	.83
Online diary—food	86.3 (822)	11 (3-28)	80.6 (307)	8 (1-25)	.009	.006
Online diary—exercise	77.7 (740)	4 (1-11)	73.2 (279)	5 (0-14)	.09	.46
Discussion forum posts	39.9 (380)	0 (0-2)	73.8 (281)	5 (0-22)	<.001	<.001
Discussion forum views	19.6 (187)	0 (0-0)	55.4 (211)	1 (0-8)	<.001	<.001
Accessed menu plan	91.0 (867)	3 (1-9)	96.3 (367)	8 (3-15)	<.001	<.001
Accessed physical activity plan	84.2 (802)	2 (1-4)	94.8 (361)	6 (3-11)	<.001	<.001
Accessed weekly educational tips and challenges	88.9 (847)	2 (1-5)	93.4 (356)	5 (2-8)	.01	<.001
Attended weekly online meeting (webinar)	N/A	N/A	57.0 (217)	1 (0-2)	N/A	N/A

[a]Frequency is the number of days the feature was used.

[b]Difference between the two programs tested using chi-square test.

[c]Difference between the two programs tested using Wilcoxon Rank Sum test.

Discussion

Principal Findings

The primary aim of this study was to determine whether an 8-week challenge version of a commercial Web-based weight loss program, which integrated persuasive features including system credibility support through the use of a celebrity personal trainer to endorse the program and host additional opportunities for social support, demonstrated greater initial weight loss and program engagement compared to the standard program. In the current study, the 8-week challenge version facilitated greater weight loss and engagement, but dropout and nonusage attrition rates were comparable to the standard program.

The true weight loss achieved by participants is likely to be in the range between the GLMM and LOCF results, that is, -3.4% to -6.0% for the SC and -2.9% to -5.0% for BLC, due to reasons previously described [12]. At the group level, the difference in weight loss between the groups was statistically different, but the effect size was small, suggesting that the difference between the two groups weight loss may not be clinically significant. Although there was no significant difference in attrition rates between the two programs, there was a trend for lower dropout attrition for SC participants. Furthermore, frequency of use of the website was superior for SC participants, and they were more likely to use some website components. SC participants engaged more with the standard social support components (eg, discussion forum) SC participants were also more likely to access educational materials than BLC.

The differences in initial weight loss and engagement between the BLC and SC programs could be partly explained by the inclusion of additional persuasive features. By offering more opportunities for social support to SC participants, a more supportive environment may have been created [13]. The support environment may have allowed SC participants to compare their performance to others (social comparison) or motivated participants to change their behavior if they recognized that other participants were successfully making change (social facilitation). Furthermore, the use of a celebrity personal trainer to endorse and provide content within the SC program may have been more persuasive by boosting the perceived credibility of the program. Alternatively, the differences in weight loss and engagement may be due to the more stringent energy intake and expenditure targets set as part of the SC program. For example, greater access to the educational materials may have been required to facilitate adherence with the targets, suggesting that these resources may be necessary to facilitate adherence to the energy intake and expenditure targets set as part of SC program. However, as compliance to the recommended energy intake and expenditure targets for each program were not measured and an observational study design was used, we cannot be sure which components of SC lead to greater engagement and weight loss (ie, the persuasive features of social support and/or system credibility support, or different energy expenditure and intake goals).This could be examined in future studies.

Limitations

The weight loss analysis used participants' weigh-in records self-reported online. However, the accuracy of weight self-reported on the Internet has been shown to be reasonable [14]. Outcomes were evaluated only during the 8-week challenge; therefore, the long-term impact on user engagement and weight loss were not considered. As the SC program was more expensive for participants to subscribe to than the BLC program, future studies should also consider the cost-effectiveness of the programs when evaluating program effectiveness. Finally, the use of a celebrity personal trainer also limits the external validity of the results.

Conclusion

This preliminary observational study supports the need for further evaluation of Web-based weight loss programs that incorporate persuasive strategies, including enhanced credibility support and social support, to enhance initial weight loss and engagement. Future randomized control trials accompanied by mediation analyses should specifically determine which intervention components (ie, persuasive features: social support and/or celebrity endorsement, or stringent eating and physical activity recommendations) of the Web-based program are associated with improvements in engagement and weight loss, and whether initial weight loss and engagement are maintained in the long term.

Conflicts of Interest

CEC has been a nutrition consultant to SP Health Co and MJH received a PhD scholarship supplement from SP Health Co and partial fellowship from the Penn Foundation. All other authors declare that they have no competing interests.

Multimedia Appendix 1

Online food and exercise diary used by SC and BLC participants.

[JPG File, 201KB - jmir_v15i7e129_app1.jpg]

Multimedia Appendix 2

Weekly self-monitoring of weight for BLC and SC.

[JPG File, 235KB - jmir_v15i7e129_app2.jpg]

Multimedia Appendix 3

Additional opportunity for social support for SC participants: Facebook page.

[JPG File, 268KB - jmir_v15i7e129_app3.jpg]

Multimedia Appendix 4

Additional opportunity for social support for SC participants: Weekly online meeting.

[JPG File, 277KB - jmir_v15i7e129_app4.jpg]

References

1. Neve M, Morgan PJ, Jones PR, Collins CE. Effectiveness of web-based interventions in achieving weight loss and weight loss maintenance in overweight and obese adults: a systematic review with meta-analysis. Obes Rev 2010 Apr;11(4):306-321. [doi: 10.1111/j.1467-789X.2009.00646.x] [Medline: 19754633]
2. Kodama S, Saito K, Tanaka S, Horikawa C, Fujiwara K, Hirasawa R, et al. Effect of Web-based lifestyle modification on weight control: a meta-analysis. Int J Obes (Lond) 2012 May;36(5):675-685. [doi: 10.1038/ijo.2011.121] [Medline: 21694698]
3. Wieland LS, Falzon L, Sciamanna CN, Trudeau KJ, Brodney S, Schwartz JE, et al. Interactive computer-based interventions for weight loss or weight maintenance in overweight or obese people. Cochrane Database Syst Rev 2012;8:CD007675. [doi: 10.1002/14651858.CD007675.pub2] [Medline: 22895964]
4. Krukowski RA, Harvey-Berino J, Bursac Z, Ashikaga T, West DS. Patterns of success: online self-monitoring in a web-based behavioral weight control program. Health Psychol 2013 Feb;32(2):164-170. [doi: 10.1037/a0028135] [Medline: 22545978]
5. Nackers LM, Ross KM, Perri MG. The association between rate of initial weight loss and long-term success in obesity treatment: does slow and steady win the race? Int J Behav Med 2010 Sep;17(3):161-167. [doi: 10.1007/s12529-010-9092-y] [Medline: 20443094]
6. Oinas-Kukkonen H, Harjumaa M. Persuasive systems design: Key issues, process model, and system features. Communications of the Association for Information Systems 2009;24(1):28.

XSL•FO
RenderX

7.   Kelders SM, Kok RN, Ossebaard HC, Van Gemert-Pijnen JE. Persuasive system design does matter: a systematic review of adherence to web-based interventions. J Med Internet Res 2012;14(6):e152 [FREE Full text] [doi: 10.2196/jmir.2104] [Medline: 23151820]

8.   Lehto T. Persuasive features in six weight loss websites. In: Persuasive Technology: 5th International Conference.: Springer; 2010 Presented at: PERSUASIVE; June 7-10, 2010; Copenhagen p. 162-173.

9.   Biggest Loser Club Australia. URL: http://www.biggestloserclub.com.au/ [accessed 2013-05-22] [WebCite Cache ID 6GnugNURa]

10. Collins CE, Morgan PJ, Jones P, Fletcher K, Martin J, Aguiar EJ, et al. A 12-week commercial web-based weight-loss program for overweight and obese adults: randomized controlled trial comparing basic versus enhanced features. J Med Internet Res 2012;14(2):e57 [FREE Full text] [doi: 10.2196/jmir.1980] [Medline: 22555246]

11. Eysenbach G. The law of attrition. J Med Internet Res 2005;7(1):e11 [FREE Full text] [doi: 10.2196/jmir.7.1.e11] [Medline: 15829473]

12. Neve M, Morgan PJ, Collins CE. Weight change in a commercial web-based weight loss program and its association with website use: cohort study. J Med Internet Res 2011;13(4):e83 [FREE Full text] [doi: 10.2196/jmir.1756] [Medline: 21993231]

13. Kalodner CR, De Lucia JL. Components of effective weight loss programs: theory, research, and practice. J Counsel Dev 1990;68(4):427-433. [doi: 10.1002/j.1556-6676.1990.tb02523.x]

14. Harvey-Berino J, Krukowski RA, Buzzell P, Ogden D, Skelly J, West DS. The accuracy of weight reported in a web-based obesity treatment program. Telemed J E Health 2011 Nov;17(9):696-699 [FREE Full text] [doi: 10.1089/tmj.2011.0032] [Medline: 21882997]

Abbreviations

BLC: Biggest Loser Club
BMI: body mass index
GLLM: Generalized Linear Mixed Model
LOCF: last observation carried forward
SC: Shannan Ponton Fast Track Challenge

Original Paper

An Internet-Based Guided Self-Help Intervention for Panic Symptoms: Randomized Controlled Trial

Wouter van Ballegooijen[1], PhD; Heleen Riper[1,2], PhD; Britt Klein[3,4,5], DPsych (Clinical); David Daniel Ebert[2,6], PhD; Jeannet Kramer[7], PhD; Peter Meulenbeek[8,9], PhD; Pim Cuijpers[1], PhD

[1]Department of Clinical Psychology and the EMGO Institute for Health and Care Research, VU University Amsterdam, Amsterdam, Netherlands

[2]Innovation Incubator, Leuphana University, Lüneburg, Germany

[3]DVC-Research Portfolio & the Faculty of Health Sciences, University of Ballarat, Ballarat, Australia

[4]Centre of Mental Health Research, The Australian National University, Canberra, Australia

[5]National eTherapy Centre, Faculty of Life and Social Sciences, Swinburne University, Melbourne, Australia

[6]Philipps University Marburg, Marburg, Germany

[7]Innovation Centre of Mental Health & Technology, Trimbos-institute, Utrecht, Netherlands

[8]GGNet (Community Mental Health Centre), Warnsveld, Netherlands

[9]Department of Psychology, Health and Technology, University of Twente, Enschede, Netherlands

Corresponding Author:
Wouter van Ballegooijen, PhD
Department of Clinical Psychology and the EMGO Institute for Health and Care Research
VU University Amsterdam
Van der Boechorststraat 1
Amsterdam, 1081 BT
Netherlands
Phone: 31 6 24112991
Fax: 31 20 598 8758
Email: w.van.ballegooijen@vu.nl

Abstract

Background: Internet-based guided self-help is efficacious for panic disorder, but it is not known whether such treatment is effective for milder panic symptoms as well.

Objective: To evaluate the effectiveness of Don't Panic Online, an Internet-based self-help course for mild panic symptoms, which is based on cognitive behavioral principles and includes guidance by email.

Methods: A pragmatic randomized controlled trial was conducted. Participants (N=126) were recruited from the general population and randomized to either the intervention group or to a waiting-list control group. Inclusion criteria were a Panic Disorder Severity Scale-Self Report (PDSS-SR) score between 5-15 and no suicide risk. Panic symptom severity was the primary outcome measure; secondary outcome measures were anxiety and depressive symptom severity. Measurements were conducted online and took place at baseline and 12 weeks after baseline (T1). At baseline, diagnoses were obtained by telephone interviews.

Results: Analyses of covariance (intention-to-treat) showed no significant differences in panic symptom reduction between groups. Completers-only analyses revealed a moderate effect size in favor of the intervention group (Cohen's d=0.73, *P*=.01). Only 27% of the intervention group finished lesson 4 or more (out of 6). Nonresponse at T1 was high for the total sample (42.1%). Diagnostic interviews showed that many participants suffered from comorbid depression and anxiety disorders.

Conclusions: The Internet-based guided self-help course appears to be ineffective for individuals with panic symptoms. However, intervention completers did derive clinical benefits from the intervention.

Trial Registration: Nederlands Trial Register: NTR1639; http://www.trialregister.nl/trialreg/admin/rctview.asp?TC=1639 (Archived by WebCite at http://www.webcitation.org/6ITZPozs9).

KEYWORDS

Introduction

Panic disorder (PD) with or without agoraphobia is a prevalent anxiety disorder associated with substantial loss of quality of life for the patient and considerable costs to society [1-4]. Subclinical PD, defined as panic symptoms that do not meet full *Diagnostic and Statistical Manual of Mental Disorders* (Fourth Edition; *DSM-IV*) criteria for PD, is just as prevalent [2,4]. Subclinical panic symptoms can develop into clinical PD and are also a predictor for the development of mental disorders other than PD, such as generalized anxiety disorder, social phobia, or major depressive disorder (MDD) [5].

For treatment, PD can be effectively treated with psychological or drug therapy [6-8]. Research indicates that it is also possible to prevent or delay the onset of clinical PD in people with subclinical panic symptoms [9,10]. A recent study showed that a group intervention involving primarily cognitive behavioral therapy effectively reduced symptoms in subclinical cases of PD, as well as in relatively mild cases [10]. This group course, called Don't Panic, could also be acceptable from a cost-effectiveness point of view [11].

Internet-based guided self-help has shown to be an efficacious treatment of PD as well, with a large effect size (Hedge's g=0.83) [12]. To date, all but 1 study [13] comparing Internet-based guided self-help for PD with a control condition have focused purely on groups with clinical PD, which commonly was also the primary diagnosis (eg, [14,15]). These studies excluded subclinical cases (eg, [14-16]). Recently, an Internet-based version of the group course Don't Panic has been developed. This intervention, Don't Panic Online, is an Internet-based self-help course with minimal guidance specifically for individuals with mild panic symptom severity. The aim was to provide an accessible, low-intensity, early intervention for panic symptoms.

The current study is a pragmatic randomized controlled trial (RCT) of the effectiveness of Don't Panic Online in reducing panic and anxiety symptoms among participants with subclinical and mild clinical PD. We postulate a difference in effect between Don't Panic Online and a waiting-list control group.

Methods

Design

We conducted a pragmatic RCT with 2 arms: (1) Internet-based guided self-help, and (2) a waiting-list control group (see subsequent description). The Medical and Ethical Committee of VU University Medical Center approved the study protocol, which is described in greater detail elsewhere [17]. This paper was written in accordance with the CONSORT-EHEALTH checklist [18], and this trial has been registered in the Netherlands Trial Register (NTR1639). The Netherlands Trial Register is part of the Dutch Cochrane Centre.

Study Population

We included participants aged 18 and older, with subclinical PD or clinical PD with relatively mild symptom severity, who had access to the Internet. Any individuals who were at risk of suicide were excluded. Subclinical or mild PD was defined as having a score ranging from 5 to 15 on the Panic Disorder Severity Scale-Self Report (PDSS-SR) [19]. These cut-off points represent slight to moderate panic symptom severity [20]. No restrictions were imposed on the use of pharmacotherapy or psychotherapy.

Sample Size

Previous RCTs of Internet-based self-help interventions for panic symptoms showed large between-group effect sizes [12]. Our aim was to recruit participants with milder symptom severity than those who took part in these studies. Therefore, our sample was expected to show a smaller decrease in panic symptoms. Based on a moderate effect size (Cohen's d [d]= 0.50), and using a 2-sided t test (alpha = .05, power 80%) to compare the PDSS-SR scores of the intervention group with those of the control group, we aimed to include 128 participants [21], with 64 in each group. Any missing values at posttreatment were imputed.

Recruitment

Participants were recruited from the general population. Most of those who applied for participation did so after reading about this study in the health section of an online newspaper. Additional online recruitment was conducted by means of a Facebook advertising campaign and by posting messages on panic-related or anxiety-related message boards. This was supplemented by offline recruitment by means of advertisements in national newspapers and articles in local newspapers. Interested individuals were directed to a study website, where they could find information about participation and a downloadable informed consent form. The application procedure involved printing and signing the informed consent form, then sending this to the research team (either as a physical document, by conventional mail, or as a scanned document attached to an email).

Randomization and Procedure

Consenting applicants were sent an email with a link to the online questionnaires. The baseline (T0) questionnaires included the screening questionnaires for inclusion. Any participants who reported severe panic symptoms or who were at risk of suicide were sent an automatic message advising them to contact their general practitioner and/or to visit a website for suicide prevention. This website [22] offers psychoeducation and a helpline by telephone or online chat [23]. Those participants who had completed T0 and who met the inclusion criteria were contacted within 2 weeks for a diagnostic interview by telephone. This interview was used to obtain a more detailed overview of the study sample, not for the purposes of inclusion or exclusion. After the interview, all participants were randomized to 1 of the 2 groups. Randomization was stratified for the presence or absence of agoraphobic symptoms (PDSS-SR item 4 score ≥2) and the use of antidepressants or sedatives. Randomization lists were generated automatically using a computer program. The T0 measurement can be considered to be double blind because the participants were not randomized until they had completed all of the questionnaires and the diagnostic interview. Blinding of the participants at

posttreatment assessment (T1) was not possible because at that stage they were aware of the nature of the group to which they had been allocated. T1 was scheduled 12 weeks after the baseline assessment. Both T0 and T1 were self-reported and were conducted through the Internet. Any participants who had not completed T0 or T1 were sent up to 3 automated reminders by email at weekly intervals.

Intervention

Don't Panic Online is a guided, Internet-based, individual, self-help course, based on cognitive behavioral therapy principles. The course consists of 6 sessions in which the participants learn to control their panic symptoms by applying various cognitive and behavioral techniques and skills. The course's content is described in more detail elsewhere [17]. A typical lesson takes approximately 30 minutes and consists of an introduction, a discussion of the previous lesson's homework, new theory, and homework for the following week. A track-and-trace system keeps a record of the dates on which participants log on and complete a lesson. The participants in the intervention group were coached by trained, Master's-level clinical psychology students. Every week, these participants received an email from their coach, asking how they were doing and whether they were experiencing any difficulty in following the program. The coaches responded to questions about the course and the associated exercises. They also gave brief replies to questions about the participant's mental health. The coaches were supervised by the first author. On average, the total time spent on each participant was 1 to 2 hours.

Participants in the control group received access to Don't Panic Online after completing the T1 measurement (12 weeks after T0). While waiting, they had access to an information website about the symptoms of panic and agoraphobia. This website included advice to contact a general practitioner in case the participant had further questions about panic symptoms and its treatment. All participants in the control group and the intervention group were free to seek any (additional) help they might require.

Instruments

The following variables were measured: demographic data, *DSM-IV* diagnosis, symptoms of anxiety and panic, depressive symptoms, and suicide risk. All variables were measured at both T0 and T1, except for demographic data, diagnosis, and suicide risk, which were only measured at T0.

The T0 measurement started with demographic questions. These included age, gender, place of birth, marital status, education level, physical health, and previous mental health diagnoses.

The Composite International Diagnostic Interview (CIDI) 12-month prevalence [24] was used to ascertain the presence or absence of PD, other anxiety disorders, and depression. A clinical diagnosis was made, not as an inclusion criterion, but to gain a more complete overview of the participants. The CIDI, which was developed by the World Health Organization, is an extensive, fully structured, diagnostic interview to assess *DSM-IV* Axis-I diagnoses [24]. The only subscales used were depression, PD, agoraphobia, generalized anxiety disorder,

social phobia, and posttraumatic stress disorder. In this study, a trained interviewer administered the CIDI by telephone.

The severity of current panic symptoms was measured using the PDSS-SR. The PDSS, which was originally designed as a face-to-face interview for both research and clinical practice [25], was adapted to be used in a patient self-report format [19]. The instrument contains 7 items that assess the severity of 7 dimensions of PD and its associated symptoms. The PDSS-SR generates a total score ranging from 0 to 28. The higher the score, the more severe the panic symptoms. The questionnaire has adequate psychometric properties when compared with the PDSS [19,26]. For the purposes of the current study, a score of less than 5 indicates that there are no clinically significant symptoms, whereas a score of more than 15 is interpreted as severe PD. Therefore, our study focused on the group with scores ranging from 5 to 15. According to the study by Furukawa et al [20], this score range identifies participants with mild to moderate panic symptoms but excludes those without panic symptoms as well as those with severe panic symptoms.

Anxiety symptoms in general were measured using the Beck Anxiety Inventory (BAI) [27]. The BAI contains 21 short questions. Convergent and divergent validity is sufficient [28,29]. The score ranges from 0 to 63. A score of 30 or more is considered to correspond to severe anxiety symptoms.

Depressive symptoms were measured using the Center for Epidemiologic Studies Depression scale (CES-D) [30]. The CES-D is a 20-item self-report questionnaire. The score of each individual item ranges from 0 to 3, whereas the total score ranges from 0 (no feelings of depression) to 60 (severe feelings of depression). Convergent validity of the online Dutch version for adults with other depressive measures is good [31]. With a cut-off score of 22 for MDD, it also has good predictive validity [31].

Suicide risk and suicidal ideation were measured using the specific section of the Mini-International Neuropsychiatric Interview (MINI) [32,33]. The MINI suicide section consists of 6 items and classifies participants into categories ranging from no suicide risk to high suicide risk. Any individuals with a moderate to high suicide risk were excluded from this study. In the current study, these items were administered online and presented as self-report items.

An indication of health care services usage during the past month was obtained using Part I of the Trimbos and Institute of Medical Technology Assessment Questionnaire on Costs Associated with Psychiatric Illness (TiC-P) [34].

Finally, the T1 battery of online questionnaires included open questions concerning the participant's subjective experience with Don't Panic Online and reasons for not finishing the program. These questions were only administered to the intervention group.

Analyses

Firstly, means and standard deviations were calculated for age and symptom severity of panic, anxiety, and depression. Any differences in symptom severity between the intervention group and control group were expressed in terms of Cohen's *d* (see

XSL•FO

RenderX

subsequent description) to give an indication of the magnitude of the difference in question.

Between-group effects at T1 were calculated using analyses of covariance (ANCOVA), controlling for pretreatment scores. Instead of F values, t values of parameter estimates are reported because only 2 groups are compared (where $t^2 = F$). Effect sizes on continuous measures were expressed in terms of Cohen's d, which was calculated by and dividing the mean difference between the 2 mean scores by the pooled standard deviation: $(\text{mean}_1 - \text{mean}_0)/\text{SD}_{\text{Pooled}}$. Effect sizes of 0.8 can be assumed to be large, whereas effect sizes of 0.5 are moderate and effect sizes of 0.2 are small [21]. Because Cohen's d does not take covariance into account, partial η^2 is also reported in this paper. It cannot be estimated which level of partial η^2 could be considered adequate because this effect size is dependent on several factors. Within-group effects were analyzed using paired-sample t tests and expressed in terms of Cohen's d in which the correlation between T0 and T1 was taken into account by applying Morris and DeShon's equation 8 [35]. Finally, the proportion of participants below the PDSS-SR cut-off points for clinical and subclinical PD was calculated for both T0 and T1. We used the cut-off points of 8 and 5, indicating clinical PD [25] and subclinical PD [20], respectively. All analyses were conducted for the full sample, for the subgroup completers, and for subgroups with and without the diagnosis of PD according to the CIDI. We maintained a 2-sided alpha of .05. For all analyses, SPSS version 17 (SPSS Inc, Chicago, IL, USA) was used.

The data were analyzed in agreement with the intention-to-treat (ITT) principle. Missing data at T1 were imputed by multiple imputation, in which all variables except for nominal variables (ie, age, education level, clinical diagnoses, and symptom severity on all measures at T0 and T1) were included as predictors. Ten datasets were generated and analyses were performed using pooled data. Compared with single imputation methods, multiple imputation generates a more conservative estimate of the sample standard error [36] and overestimation of effect sizes and P values is unlikely. For the purpose of sensitivity analysis, P values and effect sizes were also estimated by running the Expectation Maximization (EM) algorithm [37] on the missing data.

Results

Sample

Of 368 applicants who applied and sent in informed consent forms, 126 were included in the study. See Figure 1 for a flowchart and an overview of excluded applicants. The participants were primarily female (85/126, 67.5%), born in the Netherlands (115/126, 91.3%), with a mean age of 36.6 years (SD 11.4, range 18-67), and 50% had a bachelor's degree or higher (Table 1). Diagnostic interviews showed that 97 (77.0%) of the included participants met the criteria for PD with or without agoraphobia. Other *DSM-IV* anxiety disorders and MDD were also prevalent (Table 1). Five participants (4.0%) did not meet the criteria for a diagnosis of a mood or anxiety disorder. The control group had slightly higher baseline scores than the

intervention group (Table 1), but there were no to little further differences between the intervention group and control group. Details of health care services usage (eg, visits to the general practitioner) are presented in Table 2. Approximately half of the participants reported having consulted a general practitioner in the month immediately prior to the study, and one-third had seen a psychologist or psychiatrist.

Study Dropout

The posttreatment measurement was completed by 73 participants (57.9%). There was no significant difference between the measurements and characteristics of these 73 study completers and those of the 53 participants who were lost to follow-up. However, within the intervention group, study dropouts were less likely to have completed lessons 1 to 4 of the course ($\chi^2_1 = 15.1$, $P<.001$).

Intervention Adherence

Of the 63 participants in the intervention group, 60 (95%) started lesson 1, whereas 3 participants did not log in at all (Figure 1). Approximately half of the participants (31/63, 49%) completed lesson 2. Five participants (8%) finished all 6 modules of Don't Panic Online, 4 (6%) of them within the given 3-month time frame. During the trial, 3 participants (5%) reported that they experienced difficulties accessing the website. Those participants in the intervention group who completed T1 but did not complete the intervention (n=30) were asked why they dropped out. The most frequently reported reasons involved time constraints (n=13), life events (n=5), and symptoms so severe that the individual was unable to follow the program (or parts thereof) or carry out the assignments (n=5; see Table 3).

Intention-to-Treat Analyses

After multiple imputation, ANCOVAs showed no significant difference in panic symptom severity at T1 between groups as measured by the PDSS-SR ($t = -1.17$, $P=.25$, partial $\eta^2 = .023$, $d=0.30$; Table 4). The within-group difference of the intervention group was significant ($t=3.06$, $P=.007$, $d=0.62$), as was the within-group difference of the control group, albeit with a smaller effect size ($t=2.26$, $P=.03$, $d=0.40$). The mean BAI score did not differ between groups ($t = -1.71$, $P=.09$, partial $\eta^2 = .027$, $d=0.39$; Table 4). Nor were there any differences between groups in terms of depressive symptoms, as measured by the CES-D ($t = -1.56$, $P=.12$, partial $\eta^2 = .034$, $d=0.39$; Table 4).

At T1, and with missing values imputed, 24 participants (38%) in the intervention group and 13 (20%) in the control group had PDSS-SR scores of less than 5 (ie, symptom free). This difference did not reach significance ($\chi^2 = 5.7$, $P=.07$). With regard to the cut-off point of 8 (the recommended cut-off for clinical diagnosis), 28 participants (44%) in the intervention group and 22 (35%) in the control group scored below 8 at T0. At T1, 38 participants in the intervention group (60%) and 33 participants in the control group (52%) scored below 8, a nonsignificant difference ($\chi^2 = 1.3$, $P=.43$).

Sensitivity analyses with the EM algorithm gave slightly different results. There was no significant effect between groups

on the primary outcome measure (PDSS-SR: $t_{124} = -1.79$, $P=.08$, partial $\eta^2 = .025$, $d=0.34$), but the difference in BAI anxiety symptoms did reach significance ($t_{124} = -2.33$, $P=.02$) with a moderate effect size ($d=0.46$, partial $\eta^2 = .042$). CES-D depressive symptoms also differed between groups ($t_{124} = -2.69$, $P=.008$) with a moderate effect size ($d=0.47$, partial $\eta^2 = .055$).

Table 1. Baseline characteristics of participants.

Characteristics	Total sample N=126	Intervention group n=63	Control group n=63	Difference at baseline (Cohen's d)
Demographics				
Age, mean (SD)	36.6 (11.4)	36.7 (12.2)	36.4 (10.7)	
Female, n (%)	85 (67.5)	44 (69.8)	41 (65.1)	
Born in the Netherlands, n (%)	115 (91.3)	57 (90.5)	58 (92.1)	
Living alone, n (%)	50 (39.7)	23 (36.5)	27 (42.9)	
High education,[a] n (%)	63 (50.0)	30 (47.6)	33 (52.4)	
Physical health problems, n (%)	9 (7.1)	5 (7.9)	4 (6.3)	
Previously diagnosed with a mental disorder, n (%)	47 (37.3)	22 (34.9)	25 (39.7)	
Diagnoses,[b] n (%)[c]				
PD with agoraphobia	61 (49.2)	30 (47.6)	31 (49.2)	
PD without agoraphobia	36 (29.0)	17 (27.0)	19 (30.2)	
Agoraphobia without PD	17 (13.7)	10 (15.9)	7 (11.1)	
GAD	11 (8.9)	5 (7.9)	6 (9.5)	
Social phobia	78 (62.9)	39 (61.9)	39 (61.9)	
PTSD	16 (12.9)	4 (6.3)	12 (19.0)	
MDD	53 (42.7)	27 (42.9)	26 (41.3)	
Symptom severity, mean (SD)				
Panic (PDSS-SR)	8.9 (3.0)	8.8 (3.2)	9.1 (2.8)	0.12
Anxiety (BAI)	24.9 (10.8)	23.7 (10.2)	26.0 (11.3)	0.22
Depression (CES-D)	20.8 (9.0)	20.0 (9.1)	21.6 (9.0)	0.18

[a]Defined as the equivalent of a bachelor's degree or higher.

[b]Percentages add up to more than 100% due to comorbid diagnoses.

[c]Missing data of 2 participants (n=124).

Table 2. Use of care in the past month.

Care use	T0, n (%)		T1,[a] n (%)	
	Intervention group (n=63)	Control group (n=63)	Intervention group (n=16)	Control group (n=39)
Visited general practitioner	27 (43%)	31 (49%)	2 (13%)	12 (31%)
Visited psychologist or psychiatrist	23 (37%)	17 (27%)	5 (31%)	14 (36%)
Visited other professional health care giver	18 (29%)	25 (40%)	3 (19%)	14 (36%)
Used antidepressants, sedatives, or sleeping pills	20 (32%)	23 (37%)	7 (44%)	13 (33%)

[a]Differences within groups and between groups did not reach significance.

XSL•FO
RenderX

Figure 1. Flow of participants through the study.
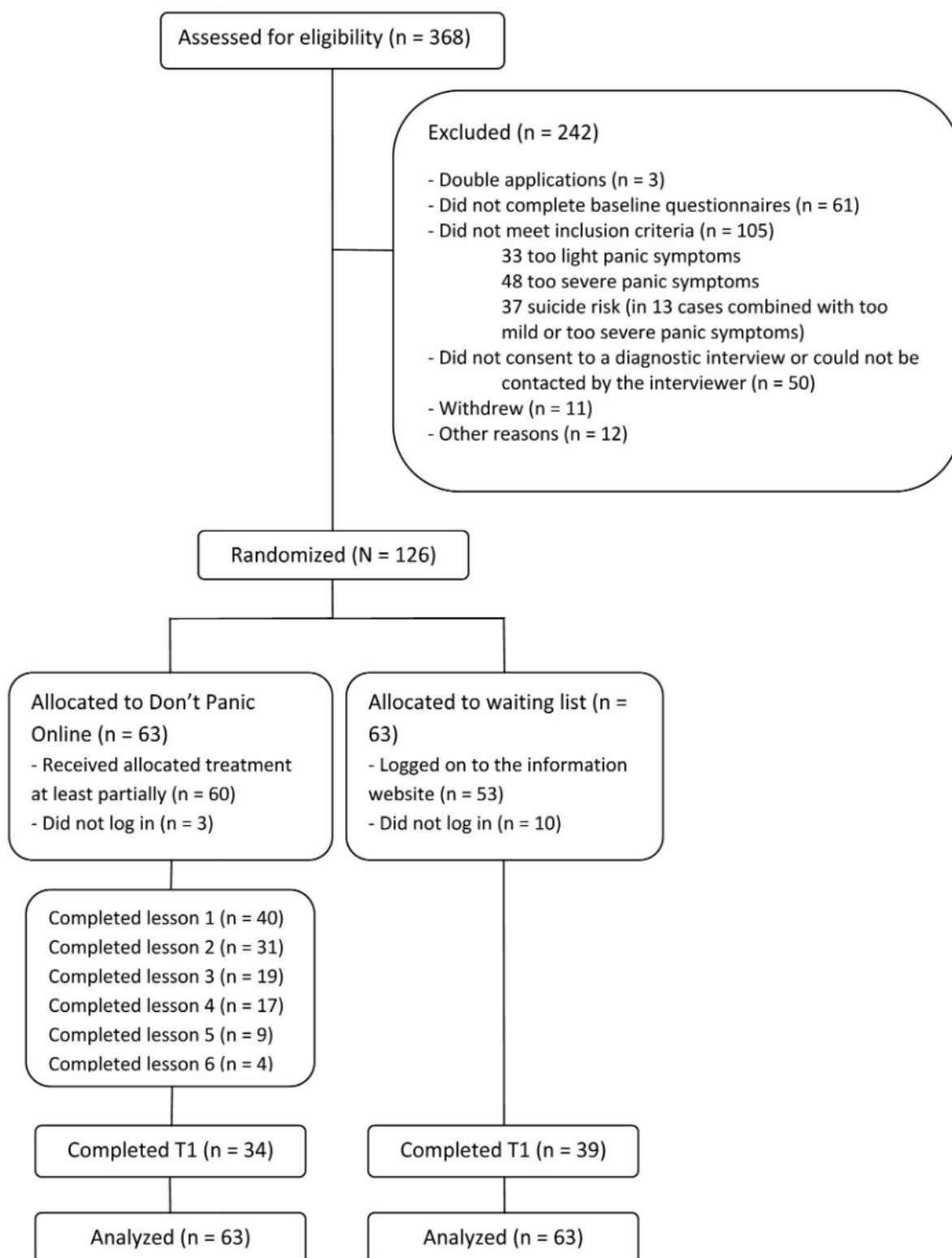
Table 3. Reasons why participants did not finish Don't Panic Online within 12 weeks (n=30).

Reason for discontinuation	n[a]
Time constraint (too busy or need more time)	13
Life events (pregnancy, loss, family issues)	5
Symptoms too severe to do assignments	5
Found other therapy	4
Content not applicable	3
Spontaneous recovery	2
Adverse effect	1
More guidance needed	1
Lack of structure	1
Lessons too slow	1
Not motivated	1

[a]Numbers do not add up to 30 because 2 participants did not give reasons and others gave several.

Table 4. Differences between groups at T1, intention-to-treat (N=126).

Measure	Group, mean (SD)[a]		Between-groups effect[a]			
	Intervention group (n=63)	Control group (n=63)	Cohen's d (95% CI)	ANCOVA[b]		
				t [c]	P	Partial η^2
PDSS-SR	5.8 (4.9)	7.3 (4.9)	0.30 (–0.91, 1.51)	–1.17	.25	.023
BAI	17.0 (12.7)	22.0 (12.7)	0.39 (–2.74, 3.53)	–1.71	.09	.027
CES-D	16.4 (12.3)	21.1 (12.1)	0.39 (–2.59, 3.42)	–1.56	.12	.034

[a]Missing data imputed by multiple imputation.

[b]Controlling for symptom severity at T0.

[c]Degrees of freedom not provided due to multiple imputation.

Completers-Only Analyses

Those participants in the intervention group who had completed the first 4 lessons (or more) of the course (n=17) were included in the completers-only analyses. These completers cannot be all considered to have completed the intervention, but after 4 lessons, participants can be considered to have experienced most of the content of the intervention. Sixteen of the 17 participants in the intervention group who had completed the first 4 lessons also filled in T1 questionnaires. Accordingly, there were 16 completers in the intervention group. These 16 individuals did not significantly differ from the noncompleters in the intervention group at T0 in terms of age, education, clinical diagnosis, and symptom severity. Control group completers were those who filled in T1 (n=39).

The ANCOVA showed significant differences between the intervention group completers and control group completers with regard to panic symptom severity at T1 ($t_{53} = -2.60$, P=.01, d=0.73; see Table 5), in favor of the intervention group. The intervention group was also characterized by a large within-group effect on panic symptoms ($t_{15} = 4.92$, P<.001,

d=1.23). In the control group, within-group effects did not reach significance. ANCOVA also showed that BAI anxiety symptom severity differed significantly between groups ($t_{53} = -2.37$, P=.02, d=0.60, see Table 5), as did depressive symptom severity, as measured using the CES-D ($t_{53} = -2.52$, P=.02, d=.94).

Ten (68%) of the intervention completers and 8 (21%) of the control group completers had a PDSS-SR score of less than 5 at T1, which is a significant difference ($\chi^2_1 = 9.1$, P=.003). In terms of the cut-off point for clinical diagnosis, 13 participants in the intervention group (81%) and 23 (59%) in the control group scored less than 8, but this difference did not reach statistical significance ($\chi^2_1 = 2.5$, P=.12).

Lastly, health care service usage rates did not differ either within or between groups (see Table 2).

Participants With Diagnosis of Panic Disorder Versus Those Without Diagnosis

Neither ITT nor completers-only analyses showed differences on any outcome measure between participants with and without clinical PD.

Table 5. Differences between groups at T1, completers[a] (n=55).

Measure	Group, mean (SD)		Between-groups effect			
	Intervention group (n=16)	Control group (n=39)	Cohen's *d* (95% CI)	ANCOVA[b]		
				t_{53}	*P*	Partial η^2
PDSS-SR	4.6 (3.3)	7.5 (4.2)	0.73 (–0.60, 2.32)	–2.60	.01	.115
BAI	15.6 (13.4)	22.6 (11.2)	0.60 (–2.93, 7.15)	–2.37	.02	.098
CES-D	12.1 (8.5)	21.6 (11.0)	0.94 (–2.50, 5.10)	–2.52	.02	.109

[a]Control group completers are those who provided posttreatment data. Intervention group completers are those who provided posttreatment data and completed at least lesson 4.

[b]Controlling for symptom severity at T0.

Discussion

Overview

This study showed that the Internet-based, guided, self-help intervention Don't Panic Online was not effective in individuals with panic symptoms. Completers-only analyses did show moderate to large effect sizes between groups in favor of the intervention group. Adherence to the treatment was low. An analysis of the data using a less conservative imputation method revealed significant effects between groups in terms of the scores for general anxiety and depressive symptoms, but not for panic symptoms. Overall, the results show that Don't Panic Online could be efficacious for intervention completers, but that it is not generally effective.

Comparison With the Literature

A meta-analysis revealed that the psychological treatment (offline and online) of full-blown PD is highly effective compared to a waiting-list control group, with a mean effect size of *d*=1.19 [8]. Samples in which more than 50% of the participants had comorbid disorders did not benefit as much, but they still showed a large effect size even when compared with pooled active and nonactive control groups (*d*=0.83) [8]. Self-help interventions have an average effect size of *d*=0.75, again when compared with pooled control groups [8]. The results of our completers-only analyses are in-line with these findings. Treatment adherence is not reported in this meta-analysis, only study dropout rates, which averaged 9.53% for intervention groups.

For study design and intervention, our study is comparable with the trial of Meulenbeek et al [10]. That study found a moderate effect size of *d*=0.68 for the face-to-face group course Don't Panic, an intervention with similar content to Don't Panic Online. Treatment completion, defined as having followed at least 6 of the 8 sessions, was 75%. In that study, the participants had a relatively low baseline mean PDSS-SR score (7.2), which is similar to our study's findings. Aside from panic symptoms, however, the sample differed from ours in a number of ways. Meulenbeek et al excluded participants with severe disorders other than PD, as well as those with social problems, and those who were receiving treatment for panic symptoms. In general, group interventions are no more effective than guided self-help interventions [38]. Possibly, any differences in outcome between

the trials of Don't Panic and Don't Panic Online might be attributed to inclusion criteria.

Previous studies that compared Internet-based guided self-help for panic symptoms with a control group showed an overall effect size of Hedge's *g*=0.83 [12]. Similar to Don't Panic Online, the interventions studied were based on cognitive behavioral therapy and were similar in length [14-16]. Compared with these studies, effect sizes in the current study were expected a priori to be lower. We included a less severe group, thereby ruling out large decreases in symptom severity. Accordingly, assuming that there was no deterioration in the control group, the difference between the intervention group and control group at T1 could not be as large. With regard to low treatment adherence, this was not found in previous studies and values ranged from 79% to 95% [12,15].

There are several differences between our study and previous studies that may have had an impact on adherence. Firstly, all participants in our trial were free to use medication and find other treatment. Some may have found other help and decided to quit Don't Panic Online. Secondly, our participants reported difficulties accessing the website. Thirdly, previous researchers had more telephone contact with their participants [14,15]. Our participants were also not interviewed after the treatment, whereas a scheduled interview after treatment may have led to better adherence [39]. Fourthly, the intervention we studied was not the same as the interventions of other studies. Perhaps Don't Panic Online is not as effective or attractive as those examined in other studies. Lastly, our sample included a large proportion of participants with comorbid disorders, and possibly a proportion of participants who did not have PD as primary diagnosis. Perhaps an Internet-based intervention specifically for panic symptoms is less suited to this group. However, epidemiological data show that panic symptoms often coincide with psychiatric disorders other than PD [2,4]. Therefore, the participants of our study appear to be a representative sample of individuals with panic symptoms.

In summary, both clinical effect and treatment adherence were lower in our study than in previous studies of Internet-based self-help interventions and the Don't Panic group course. The differences in sample characteristics between our study and previous trials could indicate that Internet-based interventions for panic symptoms are efficacious, but they may not be effective for all individuals seeking help for panic symptoms.

Limitations

When interpreting our results, several limitations should be taken into account. One limitation of this study is nonresponse at the posttreatment measurement. For a large proportion of participants, it is unknown whether their panic symptoms increased, decreased, or remained stable. These missing values were estimated by multiple imputation. Although this can be considered a conservative imputation method, it is unlikely that the imputed values greatly underestimate the intervention effect. This is because many of the participants who did not respond at T1 also left the intervention after 1 or 2 sessions, and are, therefore, unlikely to have gained much benefit from it. A second limitation is that the intervention completers are small in number and may not be representative of the intervention group as a whole, even though there did not appear to be significant differences between completers and noncompleters. The comparison of this select group with the control group, for completers-only analyses, should be interpreted with caution. Thirdly, the control group could have had gained some benefit from the information website, which could have decreased the difference between T1 mean scores of the intervention group and control group. If that is the case, our study proved that Don't Panic Online has, in general, no added value compared with an information website and our conclusion would remain the same. A fourth limitation is the lack of a follow-up measurement. It is not known whether the participants in either the intervention group or the control group showed any further improvement over the subsequent months to a year. Finally, all continuous measures were obtained by online self-report. The PDSS-SR could potentially yield lower mean scores than the PDSS interview [26], whereas online versions of questionnaires could potentially yield higher mean scores than pencil-and-paper versions [40,41]. These differences in psychometric properties limit the comparison of this study with other studies. However, this imposes no restrictions on comparisons between the intervention group and control group within our own study and, additionally, online and pencil-and-paper versions of panic questionnaires do appear to be equivalent [42,43].

Implications and Future Research

Although previous research indicates that Internet interventions can be an efficacious treatment of panic symptoms, our results may suggest that a linear program targeting specific symptoms is not always effective. As our study and others have shown, panic symptoms generally coincide with comorbid symptoms. Therefore, transdiagnostic and tailorable interventions could be a future direction of Internet-based treatment of panic. Internet-based transdiagnostic self-help programs, tailored to the anxiety and/or depressive symptoms of the participant, show promising results in terms of the treatment of panic and other common mental disorders [13,44,45]. Tailored interventions could be more effective for individuals with higher symptom severity and comorbidity rates than nontailored programs [46]. Tailoring might help to increase treatment adherence because participants would then only see those sections that are applicable to them. Given the results of our study, the further development of transdiagnostic and tailorable Internet interventions should be encouraged.

Future research could focus on identifying those groups for whom Internet-based self-help interventions are effective, for example, by means of predictor and mediator analyses. Further research is also needed to investigate ways of boosting treatment adherence to Don't Panic Online, of making it a feasible intervention for mild to moderate panic symptoms, and perhaps of modifying it to become more tailored and transdiagnostic in nature. This was the first study of Internet-based guided self-help for mild panic symptoms and our study needs to be replicated before we can draw any definitive conclusions. Lastly, although the efficacy of Internet-based guided self-help interventions has been established in several studies, it should be encouraged to conduct more pragmatic RCTs to examine the effectiveness.

Acknowledgments

Conflicts of Interest

None declared.

Multimedia Appendix 1

CONSORT-EHEALTH checklist V1.6.2 [18].

[PDF File (Adobe PDF File), 995KB - jmir_v15i7e154_app1.pdf]

References

1. Taylor CB. Panic disorder. BMJ 2006 Apr 22;332(7547):951-955 [FREE Full text] [doi: 10.1136/bmj.332.7547.951] [Medline: 16627512]
2. Batelaan N, De Graaf R, Van Balkom A, Vollebergh W, Beekman A. Thresholds for health and thresholds for illness: panic disorder versus subthreshold panic disorder. Psychol Med 2007 Feb;37(2):247-256. [doi: 10.1017/S0033291706009007] [Medline: 17076912]
3. Batelaan N, Smit F, de Graaf R, van Balkom A, Vollebergh W, Beekman A. Economic costs of full-blown and subthreshold panic disorder. J Affect Disord 2007 Dec;104(1-3):127-136. [doi: 10.1016/j.jad.2007.03.013] [Medline: 17466380]

XSL•FO

RenderX

4. Bystritsky A, Kerwin L, Niv N, Natoli JL, Abrahami N, Klap R, et al. Clinical and subthreshold panic disorder. Depress Anxiety 2010 Apr;27(4):381-389 [FREE Full text] [doi: 10.1002/da.20622] [Medline: 20013957]

5. Kinley DJ, Walker JR, Enns MW, Sareen J. Panic attacks as a risk for later psychopathology: results from a nationally representative survey. Depress Anxiety 2011 May;28(5):412-419. [doi: 10.1002/da.20809] [Medline: 21400640]

6. Furukawa TA, Watanabe N, Churchill R. Psychotherapy plus antidepressant for panic disorder with or without agoraphobia: systematic review. Br J Psychiatry 2006 Apr;188:305-312 [FREE Full text] [doi: 10.1192/bjp.188.4.305] [Medline: 16582055]

7. Schmidt NB, Keough ME. Treatment of panic. Annu Rev Clin Psychol 2010;6:241-256. [doi: 10.1146/annurev.clinpsy.121208.131317] [Medline: 20001727]

8. Sánchez-Meca J, Rosa-Alcázar AI, Marín-Martínez F, Gómez-Conesa A. Psychological treatment of panic disorder with or without agoraphobia: a meta-analysis. Clin Psychol Rev 2010 Feb;30(1):37-50. [doi: 10.1016/j.cpr.2009.08.011] [Medline: 19775792]

9. Gardenswartz CA, Craske MG. Prevention of panic disorder. Behavior Therapy 2001 Sep;32(4):725-737. [doi: 10.1016/S0005-7894(01)80017-4]

10. Meulenbeek P, Willemse G, Smit F, van Balkom A, Spinhoven P, Cuijpers P. Early intervention in panic: pragmatic randomised controlled trial. Br J Psychiatry 2010 Apr;196(4):326-331 [FREE Full text] [doi: 10.1192/bjp.bp.109.072504] [Medline: 20357312]

11. Smit F, Willemse G, Meulenbeek P, Koopmanschap M, van Balkom A, Spinhoven P, et al. Preventing panic disorder: cost-effectiveness analysis alongside a pragmatic randomised trial. Cost Eff Resour Alloc 2009;7:8 [FREE Full text] [doi: 10.1186/1478-7547-7-8] [Medline: 19393084]

12. Andrews G, Cuijpers P, Craske MG, McEvoy P, Titov N. Computer therapy for the anxiety and depressive disorders is effective, acceptable and practical health care: a meta-analysis. PLoS One 2010;5(10):e13196 [FREE Full text] [doi: 10.1371/journal.pone.0013196] [Medline: 20967242]

13. Silfvernagel K, Carlbring P, Kabo J, Edström S, Eriksson J, Månson L, et al. Individually tailored internet-based treatment for young adults and adults with panic attacks: randomized controlled trial. J Med Internet Res 2012;14(3):e65 [FREE Full text] [doi: 10.2196/jmir.1853] [Medline: 22732098]

14. Carlbring P, Bohman S, Brunt S, Buhrman M, Westling BE, Ekselius L, et al. Remote treatment of panic disorder: a randomized trial of internet-based cognitive behavior therapy supplemented with telephone calls. Am J Psychiatry 2006 Dec;163(12):2119-2125. [doi: 10.1176/appi.ajp.163.12.2119] [Medline: 17151163]

15. Klein B, Richards JC, Austin DW. Efficacy of internet therapy for panic disorder. J Behav Ther Exp Psychiatry 2006 Sep;37(3):213-238. [doi: 10.1016/j.jbtep.2005.07.001] [Medline: 16126161]

16. Wims E, Titov N, Andrews G, Choi I. Clinician-assisted Internet-based treatment is effective for panic: A randomized controlled trial. Aust N Z J Psychiatry 2010 Jul;44(7):599-607. [doi: 10.3109/00048671003614171] [Medline: 20560847]

17. van Ballegooijen W, Riper H, van Straten A, Kramer J, Conijn B, Cuijpers P. The effects of an Internet based self-help course for reducing panic symptoms--Don't Panic Online: study protocol for a randomised controlled trial. Trials 2011;12:75 [FREE Full text] [doi: 10.1186/1745-6215-12-75] [Medline: 21396089]

18. Eysenbach G, CONSORT-EHEALTH Group. 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]

19. Houck PR, Spiegel DA, Shear MK, Rucci P. Reliability of the self-report version of the panic disorder severity scale. Depress Anxiety 2002;15(4):183-185. [doi: 10.1002/da.10049] [Medline: 12112724]

20. Furukawa TA, Katherine Shear M, Barlow DH, Gorman JM, Woods SW, Money R, et al. Evidence-based guidelines for interpretation of the Panic Disorder Severity Scale. Depress Anxiety 2009;26(10):922-929 [FREE Full text] [doi: 10.1002/da.20532] [Medline: 19006198]

21. Cohen J. Statistical Power Analysis for the Behavioral Sciences. Hillsdale, NJ: Lawrence Erlbaum Associates; 1987.

22. 113 online. URL: http://www.113online.nl/ [accessed 2013-07-23] [WebCite Cache ID 6IKROX7jj]

23. Mokkenstorm JK, Huisman A, Kerkhof AJ. [Suicide prevention via the internet and the telephone: 113Online]. Tijdschr Psychiatr 2012;54(4):341-348 [FREE Full text] [Medline: 22508352]

24. Robins LN, Wing J, Wittchen HU, Helzer JE, Babor TF, Burke J, et al. The Composite International Diagnostic Interview. An epidemiologic Instrument suitable for use in conjunction with different diagnostic systems and in different cultures. Arch Gen Psychiatry 1988 Dec;45(12):1069-1077. [Medline: 2848472]

25. Shear MK, Rucci P, Williams J, Frank E, Grochocinski V, Vander Bilt J, et al. Reliability and validity of the Panic Disorder Severity Scale: replication and extension. J Psychiatr Res 2001;35(5):293-296. [Medline: 11591432]

26. Wuyek LA, Antony MM, McCabe RE. Psychometric properties of the panic disorder severity scale: clinician-administered and self-report versions. Clin Psychol Psychother 2011;18(3):234-243. [doi: 10.1002/cpp.703] [Medline: 21110405]

27. Beck AT, Epstein N, Brown G, Steer RA. An inventory for measuring clinical anxiety: psychometric properties. J Consult Clin Psychol 1988 Dec;56(6):893-897. [Medline: 3204199]

28. de Beurs E, Wilson KA, Chambless DL, Goldstein AJ, Feske U. Convergent and divergent validity of the Beck Anxiety Inventory for patients with panic disorder and agoraphobia. Depress Anxiety 1997;6(4):140-146. [Medline: 9559283]

XSL•FO
RenderX

29. Lykke J, Hesse M, Austin SF, Oestrich I. Validity of the BPRS, the BDI and the BAI in dual diagnosis patients. Addict Behav 2008 Feb;33(2):292-300. [doi: 10.1016/j.addbeh.2007.09.020] [Medline: 17976924]

30. Radloff LS. The CES-D Scale: A self-report depression scale for research in the general population. Applied Psychological Measurement 1977 Jun 01;1(3):385-401. [doi: 10.1177/014662167700100306]

31. Donker T, van Straten A, Marks I, Cuijpers P. Brief self-rated screening for depression on the Internet. J Affect Disord 2010 May;122(3):253-259. [doi: 10.1016/j.jad.2009.07.013] [Medline: 19679358]

32. Lecrubier Y, Sheehan D, Weiller E, Amorim P, Bonora I, Harnett Sheehan K, et al. The Mini International Neuropsychiatric Interview (MINI). A short diagnostic structured interview: reliability and validity according to the CIDI. European Psychiatry 1997;12(5):224-231. [doi: 10.1016/S0924-9338(97)83296-8]

33. Sheehan DV, Lecrubier Y, Sheehan KH, Amorim P, Janavs J, Weiller E, et al. The Mini-International Neuropsychiatric Interview (M.I.N.I.): the development and validation of a structured diagnostic psychiatric interview for DSM-IV and ICD-10. J Clin Psychiatry 1998;59 Suppl 20:22-33;quiz 34. [Medline: 9881538]

34. Hakkaart-van Roijen L, van Straten A, Donker M, Tiemens B. Handleiding Trimbos/iMTA questionnaire for Costs associated with Psychiatric illness (TiC-P). Rotterdam: Insituut voor Medische Technology Assessment, Erasmus MC; 2002.

35. Morris SB, DeShon RP. Combining effect size estimates in meta-analysis with repeated measures and independent-groups designs. Psychol Methods 2002 Mar;7(1):105-125. [Medline: 11928886]

36. Donders AR, van der Heijden GJ, Stijnen T, Moons KG. Review: a gentle introduction to imputation of missing values. J Clin Epidemiol 2006 Oct;59(10):1087-1091. [doi: 10.1016/j.jclinepi.2006.01.014] [Medline: 16980149]

37. Little RJA, Rubin DB. Statistical Analysis with Missing Data. Hoboken, NJ: Wiley; 2002.

38. Cuijpers P, Donker T, van Straten A, Li J, Andersson G. Is guided self-help as effective as face-to-face psychotherapy for depression and anxiety disorders? A systematic review and meta-analysis of comparative outcome studies. Psychol Med 2010 Dec;40(12):1943-1957. [doi: 10.1017/S0033291710000772] [Medline: 20406528]

39. Nordin S, Carlbring P, Cuijpers P, Andersson G. Expanding the limits of bibliotherapy for panic disorder: randomized trial of self-help without support but with a clear deadline. Behav Ther 2010 Sep;41(3):267-276. [doi: 10.1016/j.beth.2009.06.001] [Medline: 20569776]

40. Cuijpers P, Boluijt P, van Straten A. Screening of depression in adolescents through the Internet: sensitivity and specificity of two screening questionnaires. Eur Child Adolesc Psychiatry 2008 Feb;17(1):32-38. [doi: 10.1007/s00787-007-0631-2] [Medline: 17876508]

41. Donker T, van Straten A, Marks I, Cuijpers P. Quick and easy self-rating of Generalized Anxiety Disorder: validity of the Dutch web-based GAD-7, GAD-2 and GAD-SI. Psychiatry Res 2011 Jun 30;188(1):58-64. [doi: 10.1016/j.psychres.2011.01.016] [Medline: 21339006]

42. Austin DW, Carlbring P, Richards JC, Andersson G. Internet administration of three commonly used questionnaires in panic research: equivalence to paper administration in Australian and Swedish samples of people with panic disorder. International Journal of Testing 2006 Mar;6(1):25-39. [doi: 10.1207/s15327574ijt0601_2]

43. Carlbring P, Brunt S, Bohman S, Austin D, Richards J, Öst L, et al. Internet vs. paper and pencil administration of questionnaires commonly used in panic/agoraphobia research. Computers in Human Behavior 2007 May;23(3):1421-1434. [doi: 10.1016/j.chb.2005.05.002]

44. Carlbring P, Maurin L, Törngren C, Linna E, Eriksson T, Sparthan E, et al. Individually-tailored, Internet-based treatment for anxiety disorders: A randomized controlled trial. Behav Res Ther 2011 Jan;49(1):18-24. [doi: 10.1016/j.brat.2010.10.002] [Medline: 21047620]

45. Titov N, Andrews G, Johnston L, Robinson E, Spence J. Transdiagnostic Internet treatment for anxiety disorders: A randomized controlled trial. Behav Res Ther 2010 Sep;48(9):890-899. [doi: 10.1016/j.brat.2010.05.014] [Medline: 20561606]

46. Johansson R, Sjöberg E, Sjögren M, Johnsson E, Carlbring P, Andersson T, et al. Tailored vs. standardized internet-based cognitive behavior therapy for depression and comorbid symptoms: a randomized controlled trial. PLoS One 2012;7(5):e36905 [FREE Full text] [doi: 10.1371/journal.pone.0036905] [Medline: 22615841]

Abbreviations

ANCOVA: analysis of covariance
BAI: Beck Anxiety Inventory
CIDI: Composite International Diagnostic Interview
EM: Expectation Maximization
ITT: intention-to-treat
MDD: major depressive disorder
MINI: Mini-International Neuropsychiatric Interview
PD: panic disorder
PDSS-SR: Panic Disorder Severity Scale-Self Report
RCT: randomized controlled trial
T0: baseline

XSL•FO
RenderX

T1: posttreatment assessment (12 weeks after T0)

Overcoming Addictions, a Web-Based Application, and SMART Recovery, an Online and In-Person Mutual Help Group for Problem Drinkers, Part 1: Three-Month Outcomes of a Randomized Controlled Trial

Reid K Hester[1], PhD (Clin. Psycho.); Kathryn L Lenberg[2], MPH, PhD (Clin. Psycho.); William Campbell[3], MS(Psycho); Harold D Delaney[3], PhD(Psych)

[1]Behavior Therapy Associates, LLC, Research Division, Albuquerque, NM, United States
[2]Presbyterian Medical Group, Integrated Behavioral Health, Albuquerque, NM, United States
[3]Psychology Department, The University of New Mexico, Albuquerque, NM, United States

Corresponding Author:
Reid K Hester, PhD (Clin. Psycho.)
Behavior Therapy Associates, LLC
Research Division
9426 Indian School Rd NE Ste 1
Albuquerque, NM, 87111
United States
Phone: 1 505 345 6100
Fax: 1 505 345 4531
Email: reidkhester@gmail.com

Related Articles:

Letter: http://www.jmir.org/2013/8/e179/

Letter: http://www.jmir.org/2013/8/e180/

Abstract

Background: Overcoming Addictions (OA) is an abstinence-oriented, cognitive behavioral, Web application based on the program of SMART Recovery. SMART Recovery is an organization that has adapted empirically supported treatment strategies for use in a mutual help framework with in-person meetings, online meetings, a forum, and other resources.

Objective: To evaluate the effectiveness of OA and SMART Recovery (SR) with problem drinkers who were new to SMART Recovery. Our experimental hypotheses were: (1) all groups will reduce their drinking and alcohol/drug-related consequences at follow-up compared to their baseline levels, (2) the OA condition will reduce their drinking and alcohol/drug-related consequences more than the control group (SR), and (3) the OA+SR condition will reduce their drinking and alcohol/drug-related consequences more than the control group (SR only).

Methods: We recruited 189 heavy problem drinkers primarily through SMART Recovery's website and in-person meetings throughout the United States. We randomly assigned participants to (1) OA alone, (2) OA+attend SMART Recovery (SR) meetings (OA+SR), or (3) attend SR only. Baseline and follow-ups were conducted via GoToMeeting sessions with a Research Assistant (RA) and the study participant. We interviewed significant others to corroborate the participant's self-report. Primary outcome measures included percent days abstinent (PDA), mean drinks per drinking day (DDD), and alcohol/drug-related consequences.

Results: The intent-to-treat analysis of the 3-month outcomes supported the first hypothesis but not the others. Participants in all groups significantly increased their percent days abstinent from 44% to 72% ($P<.001$), decreased their mean drinks per drinking day from 8.0 to 4.6 ($P<.001$), and decreased their alcohol/drug-related problems ($P<.001$). Actual use relationships were found for the OA groups, between SR online meetings and improvement in PDA ($r=.261$, $P=.033$). In addition in the OA groups, the number of total sessions of support (including SR & other meetings, counselor visits) was significantly related to PDA ($r=.306$, $P=012$) and amount of improvement in alcohol-related problems ($r=.305$, $P=.012$). In the SR only group, the number of face-to-face

meetings was significantly related to all three dependent variables, and predicted increased PDA (r=.358, P=.003), fewer mean DDD (r=-.250, P=.039), and fewer alcohol-related problems (r=-.244, P=.045), as well as to the amount of improvement in all three of these variables. Six-month follow-ups have been completed, and the results are currently being analyzed.

Conclusions: These results support our first experimental hypothesis but not the second or third. All groups significantly increased their PDA and decreased both their mean DDD and their alcohol-related problems, which indicates that both interventions being investigated were equally effective in helping people recover from their problem drinking.

Trial Registration: Clinicaltrials.gov NCT01389297; http://clinicaltrials.gov/ct2/show/NCT01389297 (Archived by WebCite at http://www.webcitation.org/6Hh5JC7Yw).

Introduction

Online Interventions for People With Alcohol and Drug Problems

In the past decade, there has been a marked rise in the number of online resources available to individuals with alcohol and drug problems, and evidence has steadily mounted to support their use [1-3]. One frequently recognized benefit of this trend is that individuals who might not otherwise seek treatment will consider an online intervention [4]. The Internet also makes interventions available to drinkers who—whether due to physical infirmity, geographic isolation, or lack of resources—might have difficulty accessing traditional treatment services. As online interventions have become more prevalent, people have used these interventions on a scale that would overwhelm conventional resources [5].

Online interventions are used in a variety of contexts, from clinical settings to college dorms to free access on the Internet. They may be presented as stand-alone treatments, as the first step in a stepped model of care, as an adjunct to traditional care, or as a hybrid [2,5,6]. The form and content of these Web-based interventions vary widely, from simple text-based adaptations of brief screening instruments that take a minute or two to complete, to multisession, multimedia, interactive interventions that extend over several hour-long sessions [7-9].

Alternative Protocols

While the predominant paradigm for conceptualizing addictive behaviors in the United States is the 12-step model (eg, Alcoholics Anonymous, Narcotics Anonymous, etc), a significant proportion of individuals who are looking for help with their addictions reject 12-step programs for a variety of reasons [10]. At least some of these individuals are interested in viable alternative recovery options, often preferring approaches that provide them with more flexibility in how they define and address their addictive behavior(s). SMART Recovery (Self-Management And Recovery Training) [11] provides such individuals with a protocol that, like a 12-step program, employs the use of an interactive group component (either in person or through the use of Web-based chat rooms and a forum) while using the framework of the 4-point program (described below). However, SMART Recovery fundamentally differs from the 12-step model in that (1) "participants learn tools for addiction recovery based on the latest scientific research", (2) it avoids labeling (eg, "alcoholic" or "addict" unless individuals themselves accept that label), and (3) it does not conceptualize addiction as a disease per se (but is accepting of members' views of addiction as a disease) [12]. Anecdotal evidence from SMART Recovery meetings indicates that these aspects of the program draw participants to SMART Recovery (A.T. Horvath, personal communication, 12/2/08).

The Overcoming Addictions Web Application

The Overcoming Addictions Web Application (OA) is an abstinence-focused, cognitive-behavioral Web application [13] that we developed for SMART Recovery [11] that is based on its protocol. The program has parallel but separate modules for alcohol, marijuana, opioids, stimulants, and compulsive gambling. The interactive exercises in OA include tasks that focus on the 4-point program of SMART Recovery as well as additional activities to enhance motivation for change; track urges over time (with feedback); practice mindfulness exercises for preventing relapse [14], set goals, and make Change Plans [15]. Most other online interventions are brief interventions designed to increase users' motivation for change. OA is unusual in the realm of online interventions in that it focuses on the action stage of change.

To evaluate the effectiveness of OA and SMART Recovery, we conducted a randomized clinical trial (trial registration NCT01389297). Our experimental hypotheses were that (1) all groups will reduce their drinking and alcohol/drug-related consequences at follow-up compared to their baseline levels, (2) the OA condition will reduce their drinking and alcohol/drug-related consequences more than the control group (SR), and (3) the OA+SR condition will reduce their drinking and alcohol/drug-related consequences more than the control group (SR only).

Methods

Description of the Intervention: SMART Recovery

SMART Recovery's protocol for change combines motivational enhancement with cognitive-behavioral principles and strategies for behavior change. Its 4-point program focuses on (1) building and maintaining motivation, (2) dealing with urges, (3) managing thoughts, feelings, and behaviors, and (4) cultivating a lifestyle balance (of short- and long-term rewards) to prevent relapse.

SMART Recovery's program uses a common set of strategies to address all addictive behaviors. Their rationale for this is based on two aspects of addiction: (1) common etiological factors in both the development and maintenance of addictive behaviors (eg, affect regulation) [16], and (2) the broad applicability of cognitive-behavioral and motivational strategies that are supported by outcome research across addiction treatments [17]. For instance, alcohol, drugs, and compulsive behaviors like gambling produce powerfully reinforcing changes in affective states, at least on a short-term basis [18]. Identifying these immediate positive consequences is an important step in developing more adaptive alternatives.

SMART Recovery's menu of cognitive-behavioral and motivational strategies has been adapted from treatment interventions and it "evolves as scientific knowledge in addiction recovery evolves" [11]. Its elements are designed to help members address issues ranging from basic motivation for change to qualitative lifestyle changes intended to reduce the appeal of, and engagement in, harmful addictive behaviors.

SMART Recovery has a large and active online presence. In 2012, their website had, on average, 69,786 visits per month and 991 new subscribers on their online forum each month. The message boards now have over 50,000 registered users (a 130% increase in the last 2 years) (S Alwood, personal communication, 1/22/13). In addition to their online presence, they have over 800 in-person support groups worldwide [19].

Description of the Intervention: Overcoming Addictions

OA is an action stage program designed to help users learn how to achieve and maintain abstinence. It is a self-directed and interactive Web application developed to be used either as a stand-alone intervention, an adjunct to attending SMART Recovery meetings, or as an adjunct to professional therapy for addictions (see Multimedia Appendix 1). Participants could access OA anywhere or anytime they had an Internet connection. Reviewers wishing to access the program can contact the senior author for a reviewer's access login.

The OA program contains and extends the elements of the 4-point program of SMART Recovery. Prior to registering, a user can read an overview of the program and its relationship to SMART Recovery. During registration, users provide a first name, gender, email address which is also their login username, and password. Once registration is completed, the program creates a new record in its database and personalizes content for that user (eg, Welcome back, John). The user is then taken to a homepage that lists all of the program's exercises and materials that are grouped by focus. The user can access any module of the program in any order that he or she chooses (see Figure 1 for a screenshot of a user's home page).

The first module, *Getting Started* gives an overview of the program, provides a discussion of the Stages of Change [20], and suggests exercises based on the individual's perceived stage.

The second module, *Building and Maintaining Motivation for Change*, contains a values exercise, a decisional balance exercise that asks users to weigh the pros and cons of changing, and a

cost-benefit analysis exercise that is designed to elicit "change talk" from the user (see Multimedia Appendixes 2-5).

The third module, *Dealing with Urges and Cravings*, begins with a brief discussion of urges and their relationship to sobriety and lapses/relapses. It teaches users to self-monitor their urges to use, noting the date, time, intensity, and duration of the urge, the trigger to the urge, how they handled the urge, and their reactions to how they handled it. Users are able to print out a page of self-monitoring cards so that they can easily collect these data as urges happen during their day. Later, when users enter their self-monitoring data, they are provided with graphic feedback about the frequency, intensity, and duration of their urges over time. This feedback can help users see whether they're making progress in experiencing fewer urges over time. If a user is not experiencing a gradual decline in the frequency, intensity, or duration of urges over time, the program suggests they consider additional or alternative urge-coping strategies. The module also contains the urge-coping strategies recommended by SMART Recovery, empirically supported mindfulness/relaxation exercises, and a section on medications that can help reduce urges and cravings.

In addition, exercises are available to help users identify and manage the triggers that precede urges. Identifying triggers is similar to the first step in a functional analysis of drinking behaviors [21], and users are encouraged to develop plans for managing the triggers they personally identify. It is a complex module because triggers range from simple (eg, wanting to drink more with some friends than others) to complex (eg, negative mood coupled with poor coping skills). For each domain of triggers, the program presents strategies that others have found to be helpful.

The fourth module is *Self-Managing Thoughts, Behaviors, and Feelings*. There are three exercises in this module: (1) the "ABCs" of Rational Emotive Behavior Therapy (REBT) [22], (2) unconditional self-acceptance, and (3) problem solving. The ABCs of REBT section has multiple subcomponents: dysfunctional beliefs, coping statements, changing one's self-talk to change one's feelings, and the process of analyzing and correcting dysfunctional beliefs that produce negative affect [23] (see Multimedia Appendix 6).

The fifth module is *Lifestyle Balance for Preventing Relapse*. This module has five components: regaining one's health, relaxation, goal setting, social and recreational activities, and other relapse prevention strategies. The section on regaining one's health focuses on eating and sleeping well, and exercising. The section on relaxation training targets both those with high levels of trait anxiety as well as those sensitive to situation specific anxiety (eg, when experiencing urges to drink/use). The goal-setting component focuses on setting short-term goals that are specific, measurable, achievable, realistic, and timed (eg, once a day). The section on social and recreational activities helps individuals consider and sample enjoyable and rewarding prosocial activities that are compatible with their goals and values and that make a sober life more rewarding than drinking, using drugs, or engaging in other addictive behaviors. The section on relapse prevention strategies presents relapse as a learning experience (eg, the Abstinence Violation Effect [24])

and offers some additional strategies that have not been covered in the other modules.

The appearance of the site is pleasant and uncluttered. Content is delivered via text, embedded videos and audio files, links to other sites, pop-up windows, and graphic feedback charts. The site is structured in the hybrid style, meaning that all content is available from a central matrix homepage. Once users choose a content area, their exploration of the content is constrained by tunnels that direct them through the various exercises. At the conclusion of an exercise, users have the option of continuing to the next recommended activity, or they may return to the homepage.

Like most computer-delivered interventions, users are free to access as much program content, in any order, and whenever they choose. Their use is supported by a customizable SMS (short message service) text messaging and email system that prompts them to log onto the program, reminds them of their plans for managing triggers, reiterates their reasons for staying sober, or presents motivational thoughts. These personalized messages can be delivered daily at user-defined times.

Figure 1. Overcoming Addictions Web app home page.

Experimental Design

Recruitment

Participants were recruited through a home page announcement on SMART Recovery's website, announcements of the study at SMART Recovery face-to-face and online group meetings nationally, and on their blog. We also placed a thread on the SMART Recovery online forum announcing the study and invited individuals who were new to SMART Recovery to participate in the study.

Inclusion Criteria

Criteria were (1) a minimum age of 18, (2) drinking 5 or more, or 4 or more for women, standard drinks on at least one occasion in the last 90 days, (3) have an Alcohol User Disorders Identification Test (AUDIT, [25]) score of 8 or higher, (4) new to Smart Recovery (ie, are just joining or have joined within the last 4 weeks), (5) have a computer at home with Internet access, and (6) have a primary treatment goal to abstain from drinking.

Exclusion Criteria

Criteria were (1) court-mandated DWI offenders, (2) a current diagnosis of drug dependence or consider themselves to be drug dependent, (3) a reported diagnostic history of psychosis or bipolar disorder not medically managed, (4) exhibit evidence of significant cognitive impairment from brain dysfunction (based on self-report and research assistant's clinical judgment during screening), (5) have an English reading level below the 8th grade, (6) are unwilling or unable to be available for follow-up appointments at 3 and 6 months from enrollment into the study, and (7) unwilling or unable to provide one Significant Other (SO) for corroboration of participant's self-reported drinking and drug use (if any).

A minimum AUDIT score of 8 suggests that the person is at least "at risk" for alcohol-related problems. It is important to recruit participants who are new to SMART Recovery to evaluate its *initial* effect on their drinking, drug use, and related consequences. A computer with Internet access at home is necessary for participants to use the Web application.

Regarding exclusion criteria, court-mandated DWI offenders are often required to attend self-help groups, and we were concerned that these treatment-mandated offenders would have no motivation to continue beyond their mandated participation. Furthermore, such a group could prove to be difficult to find at follow-up assessments. Since the primary focus is on drinking, those with either a current diagnosis of drug dependence *or those who consider themselves to be drug dependent* were excluded. Criteria 3 and 4 reflect the need for study participants who can reason, recall, and comprehend information both in the experimental and control group. The reading level of the OA Web application is set at an 8th grade level. Potential participants were asked about their educational level to ensure they would understand the material presented. Last, we contacted participants' SOs both to corroborate their self-report of their drinking and drug use as well as to provide them with resources that may be helpful to them in supporting their loved one's changes.

Screening

Potential participants were screened over the phone using a questionnaire addressing the inclusion criteria 1 and 4-6 and exclusion criteria questions 1-7. The research assistant administered the AUDIT over the phone and asked two quantity/frequency questions, "How often have you had 5 or more (4 or more for women) standard drinks (explained briefly) in the last 90 days?" and "During the last 90 days, have you drank as often as once a month?" A response of one or more times to both questions was sufficient to be included in the study. These two screening questions were adapted from those used by Cherpitel [26], who found them sensitive and specific in screening for alcohol abuse and dependence. We also included a question regarding suicidal thoughts, intent, or behaviors. If a participant endorsed this item, we discussed ways to access support (eg, National Suicide Hotline).

We emailed potential participants a demographic form, a patient locator form, a copy of the Brief Symptom Inventory (BSI) [27], and an Informed Consent form. BSI scores were reviewed prior to enrolling potential participants in the study; if their scores were elevated and the participant reported significant levels of distress, they were encouraged to access professional support [28]. Potential participants who screened positive, had a consenting SO, and signed the Informed Consent form were randomized to either the experimental or the control groups. The timeline for the post-baseline assessments began when the participant completed his or her baseline interview.

Randomization

We used a computer-generated stratification process for randomization. Participants were classified into blocks based on gender and ethnicity (white, hispanic, or other). Within each block, participants were randomly assigned to one of the three groups. After the first 3 months, we stopped randomizing participants to the OA only group, and we started encouraging those who had been assigned to this group to attend SR meetings. We did this because recruitment was slow and feedback from referral sources at SMART Recovery indicated that many potential participants were unwilling to be randomized to a condition that asked them to not attend SR meetings.

Assessments

We used the Timeline Follow-Back (TLFB) [29,30] to measure quantity/frequency of alcohol, drug, and tobacco use. The 90-day TLFB was administered at baseline and again at 3- and 6-month post-baseline, which provided continuous data for a total of 9 months. The TLFB was also used to collect data on study participants' attendance at SMART Recovery meetings and other recovery oriented activities in which they may have engaged. We used the Inventory of Drug Use Consequences (InDUC) to measure both lifetime and recent (last 3 months) alcohol- and drug-related consequences. The psychometric properties are described in the manual for the Drinker's Inventory of Consequences (DrInC) that was developed for Project MATCH [31]. The InDUC includes 5 subscales measuring interpersonal, intrapersonal, and physical consequences, impulse control, and social responsibility.

Baseline Interview

After participants completed and returned the consent form, BSI, Participant Locator, and demographics forms, they were scheduled for a baseline interview. We used the GoToMeeting website to complete the interview. This program allows sharing of the interviewer's screen so the assessment can be viewed by both parties. Participants used the TLFB calendar generated to prompt recall of their prior 3 months of drinking as the RA entered their data in a Web application that we developed for collecting data for this study, the Drinker's Evaluation. Participants then were guided to the InDUC and asked to complete it. At the completion of the interview, they were randomly assigned to a group. Participants and research staff were not blinded to group allocation.

Participants often wanted to discuss their histories and current struggles during the assessments. In order to limit the effect of the assessment interaction, RAs responded empathically but as briefly as possible, without soliciting further processing by the participant. Further, RAs directed, as indicated, that the participant seek help from the interventions being tested in the trial. All participants received a welcome email to the study. For those assigned to the OA conditions, there was a link to the OA registration page. For those assigned to meetings, a link to the SMART Recovery website was provided to facilitate locating available meetings.

Treatment Exposure and Fidelity

Treatment fidelity in the Web application is maintained by the nature of the technology used. All participants in the group who used the OA Web application were exposed to the same program. However, because participants were able to use the OA program and any module in it as often as they chose, the amount of exposure to the intervention, the number of modules used, and the way in which modules were used varied from participant to participant. Further, there was no a priori minimum number of sessions or modules a participant must have completed to be considered to have received the intervention. Further analysis of participants' engagement with the intervention and correlations with treatment outcome will be reported in Part 2, which will include 6-month outcomes.

Fidelity in the SR meetings and online resources also varied in two ways. First, the SMART Recovery website underwent substantial improvements in content and navigation during the course of the trial and the availability of face-to-face and online meetings increased. Second, just as with the OA app, participants decided how much or how little to avail themselves of these resources.

Maximizing Compliance With Protocols

The OA program has an integrated email feature that contacts users who have not logged into the program in a week. A personalized email encourages participants to log in and resume their progress through the program. There was no protocol for encouraging participants to attend their SMART Recovery meetings.

This study was approved by the Presbyterian Health care Services Institutional Review Board. Consent was obtained by emailing consent forms and asking for participant signature and witness signature. The consent outlined the nature and extent of participation in the trial. Participants were reminded their participation was voluntary, and they could withdraw from the study at any time. In addition, participants were told they would not be identified to anyone outside of the study staff at any time for any reason. Participants returned the consent forms via mail or scanned the documents in and emailed them.

Statistical Methods

Consistent with intent-to-treat analyses, we examined the entire sample as well as examined changes within the randomly assigned groups, both with and without imputation to account for missing data. In addition, we formed groups based on their use of either SMART Recovery meetings or the OA application to examine actual use outcomes.

Results

Figure 2 illustrates the flow of participants through the study. Approximately 358 people new to SMART Recovery contacted us and expressed interest in participating in the study. Of these, 345 participants completed an initial screen with research staff. Of these, 99 were not interested, 19 did not meet the inclusion criteria, and 38 were excluded. The initial screening forms were emailed to potential participants and returned either via fax or scanned and emailed. In total, 195 participants completed the initial consent process, submitted their completed forms, and were scheduled for an initial assessment. Of these, 189 completed the initial assessment and were randomly assigned to one of three groups. One participant requested all data be removed from the study 24 hours after completing the initial interview, and we granted the request. Nineteen participants were assigned to the OA only condition, 83 were assigned to the OA plus SMART Meetings condition (OA+SR), and 87 were assigned to the SMART Meeting only condition (SR) for a total n=189.

Recruitment began September 12, 2011 (3 pilot participants were recruited in the first 2 weeks of the study), and ended August 1, 2012. Three-month follow-ups were completed November 1, 2012. Six-month follow-ups were completed March 14, 2013.

Table 1 presents the general characteristics of the participants as a whole and by group assignment. There are several striking aspects of this group of participants. First, 60.6% (114) were female. Second, the mean education level was 16 years (SD 2.4) indicating this population generally had a college education. Third, the mean AUDIT score of 24.7 (SD 8.1) is in the high range and indicates that this group would be recommended for a more extensive diagnostic evaluation for alcohol dependence. In addition, the mean score for the BSI for men was 15.62 (SD 11.4) and for women was 18.54 (SD 13.7) suggesting that many of the participants were experiencing psychological distress when they completed the initial interview. There were no significant differences between groups on any variable.

Of the 189 participants who completed random assignment and baseline interviews, 151 (83%) completed the 3-month interview. Of the 37 for whom we do not have 3-month

follow-up data, 10 withdrew from the study, and 27 were lost to follow-up. Of the 151 with 3-month follow-up data, 83 were assigned to the OA and OA+SR groups, and 68 were assigned to the SR group.

Table 1. Pretreatment characteristics of participants by group.

Variable		Overall	Group		
			SR, n=86	SR+OA, n=83	OA, n=19
Female, n (%)		114 (60.6)	52 (61)	50 (60)	12 (63)
Age, M (SD)		44.3 (10.9)	43.4 (10.6)	44.6 (11.1)	48.3 (8.4)
Ethnicity, n (%)					
	White	170 (90.4)	76 (88.4)	77 (92.8)	17 (89.5)
	Hispanic	5 (2.7)	3 (3.5)	1 (1.2)	1 (5.3)
	Other	7 (6.9)	7 (8.1)	5 (6.0)	1 (5.3)
Education, M (SD)		16.1 (2.4)	15.93 (2.5)	16.0 (2.3)	17.3 (2.1)
AUDIT[a], M (SD)		24.7 (8.1)	24.8 (8.1)	23.95 (8.2)	27.4 (7.2)
BSI[b], M (SD)		17.4 (12.9)	19.35 (12.5)	15.95 (13.6)	14.8 (11.0)
InDUC[c], M (SD)		41.4 (17.9)	42.2 (19.0)	40.6 (17.5)	40.8 (15.6)

[a]Alcohol User Disorders Identification Test.

[b]Brief Symptom Inventory.

[c]Inventory of Drug Use Consequences.

Lost to Follow-Up

We compared baseline characteristics between those completing the 3-month follow-up and those who were lost to follow-up. No differences existed between those followed up and those lost to follow-up on the following continuous variables at baseline: age, mean drinks per drinking day, AUDIT, BSI total, InDUC recent score, or PDA. No differences across groups existed on the categorical variables of group assignment, gender, or ethnicity. Only education level demonstrated a significant difference, with those who were contacted at 3 months reporting having completing more years of education (16.3) than those lost to follow-up (15.3), t_{186}=2.20, P=.029.

Intent-to-Treat Analysis

Separate repeated measures analyses of variance were conducted to assess for significance of the change over time. Our three outcome measures were Percent Days Abstinent (PDA), Mean Standard Drinks per Drinking Day (DDD), and the InDUC Recent Total score (InDUC). Improvement over all groups from baseline to 3 months was highly significant on all three dependent variables: PDA, $F_{1,149}$=160.93, P<.001, with the mean PDA increasing from 44% to 72%; DDD, $F_{1,149}$=61.73, P<.001, with the mean decreasing from 8.0 to 4.6; and InDUC, $F_{1,149}$=122.28, P<.001, with the overall mean decreasing from 40.8 to 19.5. However, none of the tests of group differences in change over time approached significance, F≤1.0. The within-group effect sizes (Cohen's d) are presented in Table 2. Tests of effects of treatment group were carried out both as tests of Group x Time in a repeated measures approach and as ANCOVAs. None of the tests that would have been indicative of differential treatment effects approached significance.

In addition to these primary analyses conducted on participants having follow-up data, data were reanalyzed after values were imputed for participants having missing data using predictive mean matching [32], and results were essentially unchanged. That is, tests of time were again highly significant, and tests of treatment x time did not approach significance.

Actual Use Analysis

Because study participants could use these resources as much or as little as they chose to, we examined changes over time and treatment group effects for those actually using the resources of the assigned treatment, and examined relationships between engagement (eg, logging into OA, attending SR and other meetings, and counselor visits) and outcomes.

Time and Treatment Group Effects for Those Actually Treated

Although it was unclear what criterion to use to consider a participant treated, 59 (71%) of the 83 OA+SR participants completing the 3-month follow-up had completed 2 or more OA sessions, and 58 (85%) of the 68 SR participants completing the 3-month follow-up had attended 2 or more SR meetings. Using these definitions of being actually treated, improvement of treated participants over all groups from baseline to 3 months was highly significant on all three dependent variables: PDA, $F_{1,115}$=139.71, P<.001, with the mean PDA increasing from 44% to 73%; DDD, $F_{1,115}$=55.04, P<.001, with the mean decreasing from 8.3 to 4.4; and InDUC, $F_{1,115}$= 93.95, P<.001, with the overall mean decreasing from 39.6 to 18.7. However, none of the tests of group differences in change over time approached significance, P>.10.

Figure 2. Flowchart of participant flow and follow-up.
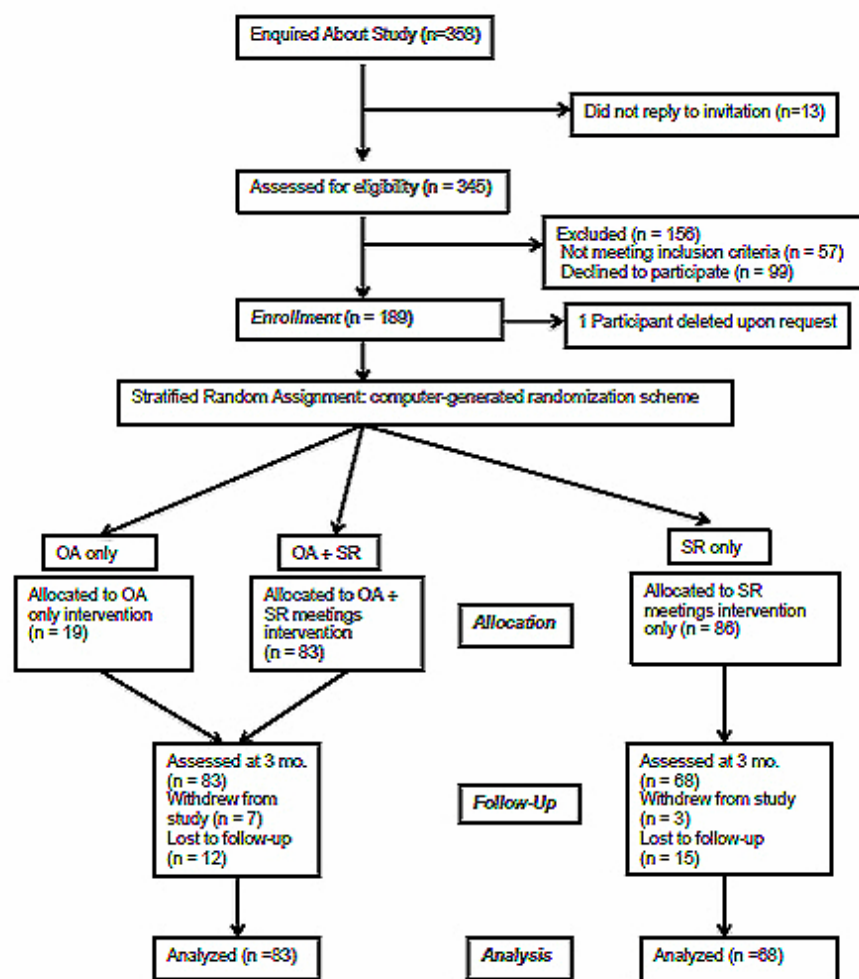
Table 2. Means and within group effect sizes for each outcome variable.

Variable & Group		Baseline	3-month follow-up	Improvement	Within group effect size d^a
Percent days abstinent (PDA)					
	OA+SR	43.83	73.32	29.49	1.00
	SR only	43.61	71.18	27.57	.91
Std. drinks per drinking day (DDD) [b]					
	OA+SR	7.88	4.59	3.29	.77
	SR only	8.25	4.66	3.59	.78
InDUC recent score [c]					
	OA+SR	40.19	19.96	20.23	1.13
	SR only	41.47	18.88	22.59	1.19

[a]Cohen's *d*.

[b]Standard drink is equal to 12 oz (355 mL) of 5% beer, 5 oz (149 mL) of 12% wine, or 1.5 oz (44 mL) of 80 proof liquor.

[c]Alcohol-related problems.

Comparisons of Those Using Only OA With Other Groups

Although we had to abandon our initial design, which included a group that would have used only OA without having the option of participating in any SR meetings, there were 29 of the 83 participants in the OA conditions who did not take part in SR meetings. This allowed post hoc comparisons to be made among three groups: those using only the OA app (n=29), those who both used the OA app and attended SR meetings (n=54), and

those randomly assigned to SR only. These three groups did not differ significantly in composition by gender, ethnicity, age, or education. Although there were no significant differences in mean baseline values on our three primary dependent variables, the trend in each case was for those in the OA only group to be more impaired initially than those who attended SR meetings. Repeated measures ANOVAs again indicated highly significant changes over time on all three dependent variables ($P<.001$), but, more importantly, tests of the group x time interaction were nonsignificant. As suggested by the plots of means in Figures 3-5, the test for differential change across the three groups did not approach significance for DDD, $F_{1,141}=0.09$, $P=.919$, or for InDUC, $F_{1,141}=0.34$, $P=.713$. For PDA, while the omnibus test of the group x time interaction was nonsignificant, $F_{1,141}=2.04$, $P=.134$, the plot of means revealed more separation of the groups. In fact, the main effect of groups on PDA was significant, $F_{2,141}=3.10$, $P=.048$, because the overall mean PDA in the OA+SR group (63.4) was greater than the average of the other two groups (53.5), $F_{1,141}=4.65$, $P=.033$. However, this resulted in part from the higher mean PDA at baseline in the OA+SR group, because there was not significant evidence of differential improvement across groups. That is, tests of interaction contrasts indicated that not only was the improvement in the SR only group (27.6) not different from that in the OA only group (23.2), $F_{1,141}=0.51$, $P=.475$, but the improvement in the OA+SR group (32.9) was also not significantly larger than the average improvement of the other two groups (25.4), $F_{1,141}=2.41$, $P=.122$.

SR Meetings or Other Support

Was the number of SR meetings, other meetings, and counselor visits predictive of the 3-month outcomes or of the improvement from baseline to 3 months for participants in the two groups? There was evidence of this, with the evidence being stronger in the SR only group than in the OA+SR condition.

Although the trend was for the SR only group to have more days of face-to-face meetings (3.31), more days of SR online meetings (5.90), and more days of Any Support (14.85) than the combined OA group (1.82, 4.42, and 12.80, respectively), these were not significantly different across conditions. For the SR only condition, the number of days of face-to-face meetings reported at 3 months was significantly predictive of all 6 of these outcome measures: PDA at 3 months ($r=.358$, $P=.003$), mean DDD ($r=-.250$, $P=.039$), and InDUC Recent Total at 3 months ($r=-.244$, $P=.045$), as well as improvement in PDA ($r=.274$, $P=.024$), mean DDD ($r=.478$, $P<.001$), and improvement in InDUC Recent Total ($r=.403$, $P=.001$). On the other hand, for this group, number of days of SR online meetings was *positively* related to mean DDD at 3 months ($r=.260$, $P=.032$). Number of days of any support for the SR group was positively related to PDA at 3 months ($r=.260$, $P=.032$) as well as to improvement in PDA ($r=.304$, $P=.012$).

In the OA+SR group (ie, excluding the 16 participants assigned to the OA only condition), neither days of face-to-face meetings nor days of SR online meetings were significantly related to any of the outcomes at 3 months or to improvement in those variables from baseline. The variable most predictive of outcomes for this group was the number of days of any support, which was significantly related to PDA at 3 months ($r=.306$, $P=.012$) and to improvement in InDUC Recent Total ($r=.305$, $P=.012$). In addition, number of days of SR online meetings was predictive of improvement in PDA ($r=.261$, $P=.033$). Relevant to the anomalous finding of the *positive* correlation in the SR group between SR online meetings and mean DDD, the correlation between these variables in the OA+SR group was slightly negative using all 67 subjects ($r=-.055$). However, if the one subject in this group who reported 83 days of online SR meetings were excluded, the correlation between number of SR meeting days and mean DDD would have been positive in this group as well ($r=.112$, $P=.372$).

Number of OA Sessions

The OA sessions completed variable was available only for those participants in the OA conditions. Participants logged into the OA program, on average, 7.2 times (SD 6.4). To assess whether there was evidence for an engagement-response relationship the number of sessions completed in the first 90 days was correlated with the values of the primary outcome variables at 3 months and with the improvement in those variables from baseline to 3 months. As shown in Table 3 below, none of these six correlations was significant. Number of days of SR online meetings was significantly predictive of improvement in PDA for the OA participants ($P=.025$). Furthermore, number of days of any support was significantly correlated with PDA at 3 months ($P=.006$), and with improvement in InDUC Recent Total ($P=.045$).

Corroboration of Self-Report Drinking by Significant Others

We collected data from 147 significant others (SO) for baseline and 3-month follow-up. In short, the reports of the SOs mirrored those given by participants. Examining the effects of time and treatment, similar to the participants, the SO data demonstrated a highly significant effect of time and nonsignificance for treatment x time effect. For mean DDD, the test of time yielded $F_{1,145}=105.25$, $P<.001$, whereas the time x treatment interaction was nonsignificant, $F_{1,145}=0.32$, $P=.573$. For PDA, the test of time yielded $F_{1,145}=140.45$, $P<.001$, whereas the time x treatment interaction was nonsignificant, $F_{1,145}=1.42$, $P=.236$.

Correlations between participants' self-report and SOs' reports were consistently highly significant: mean DDD at baseline, $r=.651$, $P<.001$; PDA at baseline, $r=.654$, $P<.001$; mean DDD at 3 months, $r=.662$, $P<.001$; PDA at 3 months, $r=.519$, $P<.001$.

Figure 3. Actual use groups: Percent days abstinent.

	Baseline	3 months
OA only	36.46	59.24
OA + SR	46.63	82.47
SR only	41.72	71.5
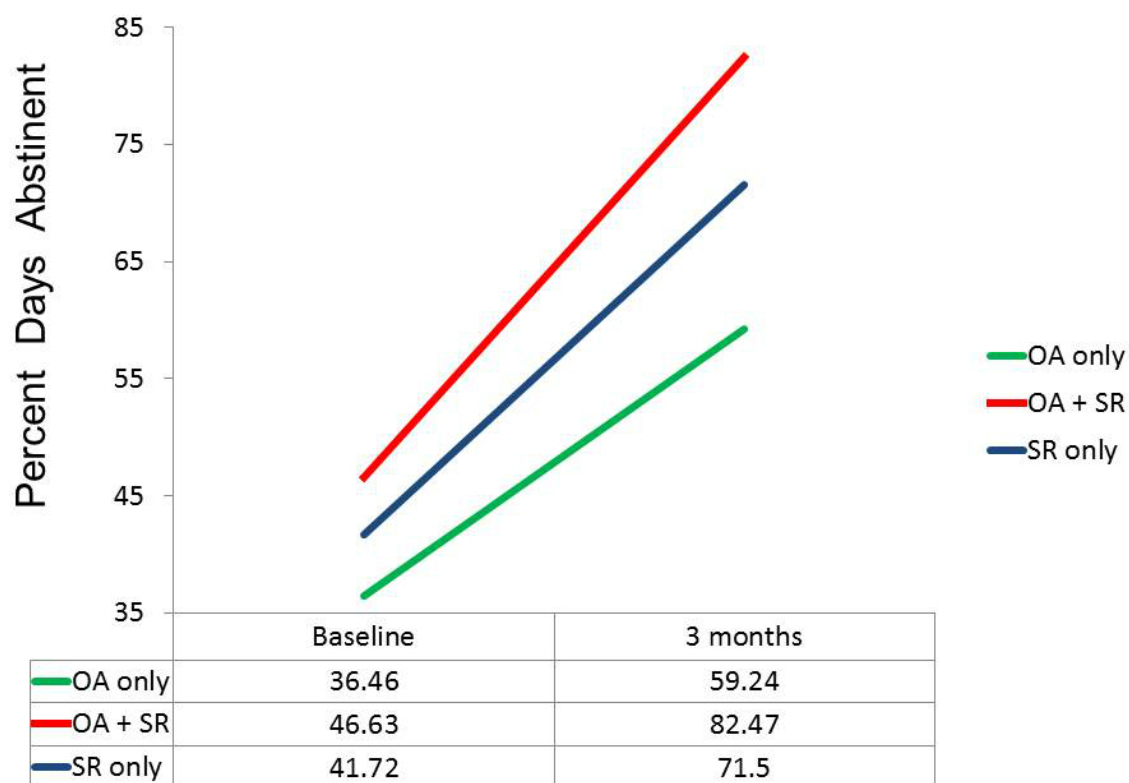
Figure 4. Actual use groups: Mean standard drinks per drinking day.

	Baseline	3 months
OA only	8.1	4.7
OA + SR	7.8	4.5
SR only	8.2	4.7

XSL•FO
RenderX
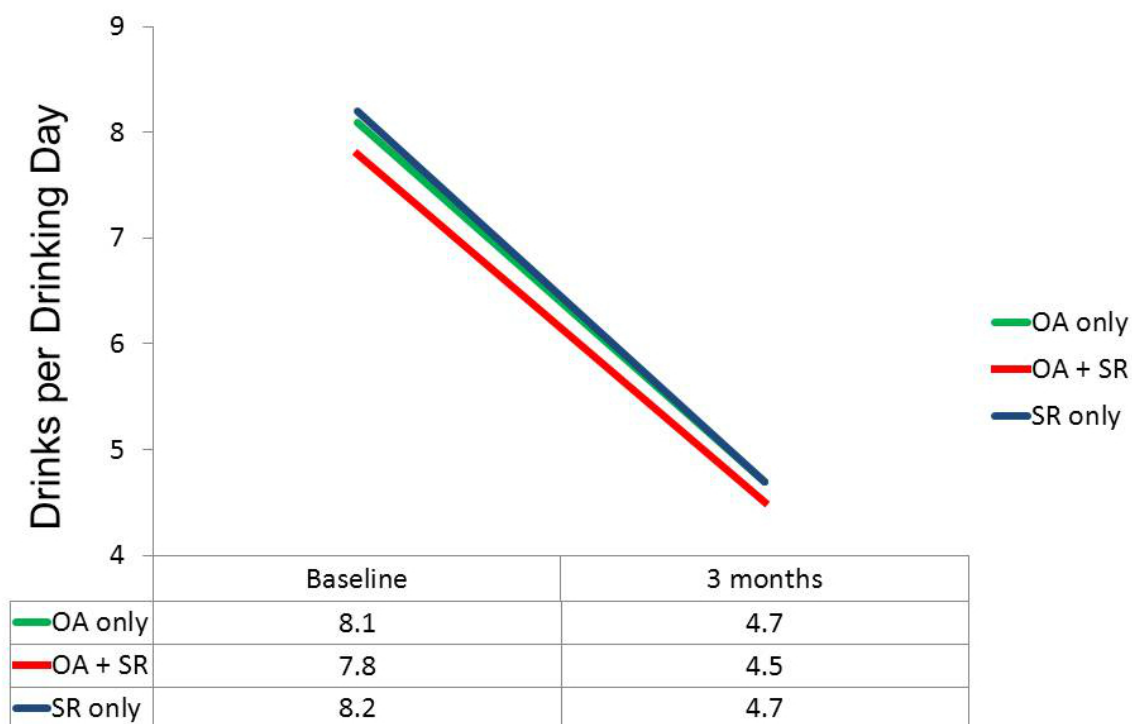
Figure 5. Actual use groups: Alcohol-related problems.
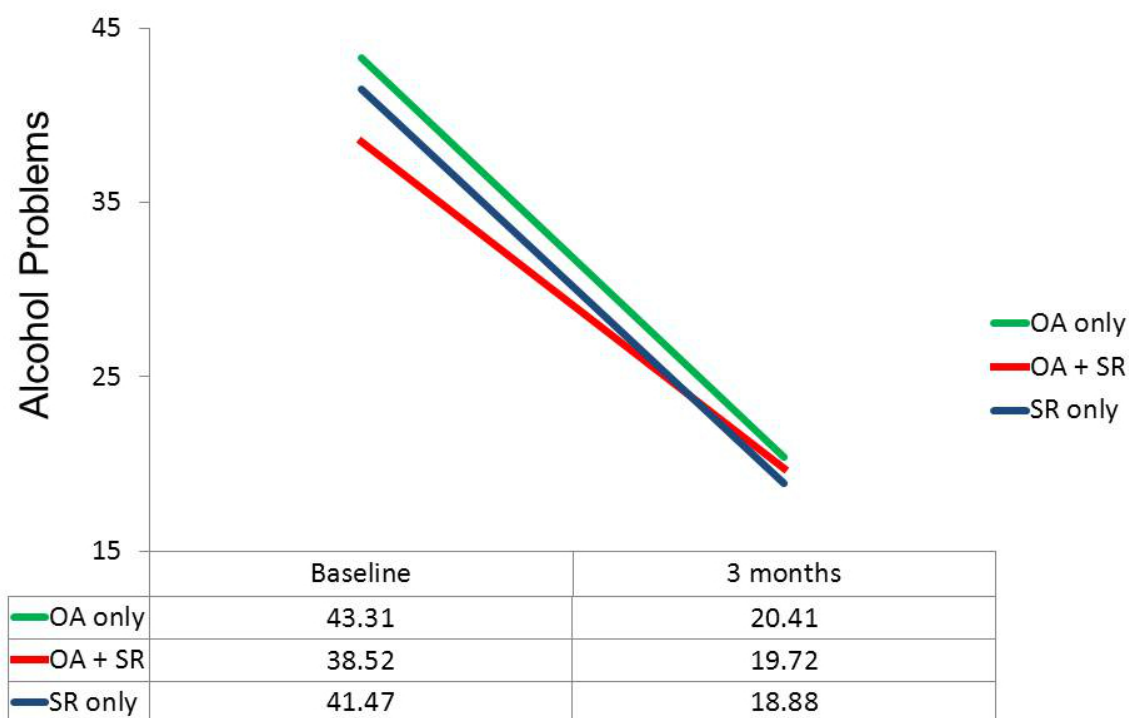
Table 3. Correlations between various resources available and outcome variables at 3 months, significant relationships noted.

Variable	PDA	DDD	InDUC	PDA Improvement	DDD Improvement	InDUC Improvement
# OA sessions baseline to 90 days	-.003	-.041	.067	-.074	.073	-.212
# SR face-to-face meetings	.138	.067	-.023	.146	.020	.074
# SR online meetings	.185	-.055	-.170	.246[a]	.079	.151
# of counselor visits	.201	-.098	.025	-.070	.031	-.019
# of other meetings	.140	-.099	-.066	.035	-.037	.170
Total of any support	.298[a]	-.115	-.148	.187	.036	.220[b]

[a]$P<.01$.

[b]$P<.05$.

Discussion

Principal Results

The experimental hypotheses were that (1) all groups will reduce their drinking and alcohol/drug-related consequences at follow-up compared to their baseline levels, (2) the OA condition will reduce their drinking and alcohol/drug-related consequences more than the control group (SR), and (3) the OA+SR condition will reduce their drinking and alcohol/drug-related consequences more than the control group (SR only). These results support our first experimental hypothesis but not the second or third.

All participants in this randomized clinical trial improved on outcomes that are important to recovery from problem drinking. Participants significantly increased their percent days abstinent per week, significantly reduced the number of drinks they

consumed on the days when they did drink, and the number of alcohol-related problems. The mean effect sizes of reductions in drinking and alcohol-related problems, averaging across the three dependent variables, were 0.97 for the OA+SR group and 0.96 for the SR only group, both being in the large range (0.8+). These statistically significant results are clinically significant. We also consider it remarkable that participants with this degree of heavy drinking made these changes over the period of 3 months.

The mean reduction in alcohol-related problems was more than 50%. While there are no norms yet for the InDUC, we have norms for the DrInC from our online Drinker's Check-up [33,34]. The only difference between the two instruments is that the InDUC adds the words "or drugs" to the questions. Since the level of drug use in the participants in this study was low (only 25% reported any drug use at baseline and the frequency of drug use in the period had a mean of 0.3 instances

and the maximum number of instances of use for any participant was 3 in 90 days), we can assume some comparability between the InDUC and DrInC scores. Assuming this comparability, participants went from the 82nd percentile at baseline to the 50th percentile at follow-up.

The correlations between attendance at SR meetings, other meetings, and counselor visits and outcomes are consistent with the perspective that the more assistance participants availed themselves of, the better their outcomes.

The analyses of how participants actually engaged with these resources present a similar picture. Significant improvements were seen on all outcome measures and no significant differences between those who only used the OA app, those who only attended meetings, and those who used both resources. The trend towards greater improvement in PDA in the group that used both resources (OA+SR) seems to be due in part to a higher level of abstinence at baseline. Conversely, the OA app only group had the lowest level of abstinence at baseline. This begs the question of whether there were other differences in this group that led them to not attend meetings. We can only speculate at this point that perhaps this group had a higher level of anxiety that may have led them to avoid attending meetings where the social norm is accountability and self-disclosure. We plan to examine this in subsequent analyses.

Attending SMART Recovery meetings appears to "work" as well as the Overcoming Addictions Web app (which is based on SMART Recovery). The reverse is also true. Having these two different ways to deliver the SMART Recovery protocol gives problem drinkers options with regards to how they learn to achieve and maintain abstinence. Some participants in our study preferred using the Web app alone. Others preferred to attend meetings. This is likely to be the case with people coming to the SMART Recovery website for the first time and considering their options. Having both protocols with equal effectiveness available increases the chances that individuals can find a path to recovery that suits them. It also increases opportunities for problem drinkers who may have limited geographical access to a face-to-face mutual support group and to those who are not inclined to attend group support meetings.

Comparison With Prior Work

The lack of differences between assigned groups in either the intent-to-treat analyses or the actual use analyses was surprising from the traditional perspective that more intervention results in better outcomes. While this is often the case in addictions

treatment outcome research, it is not always the case with freestanding online interventions. In our previous randomized clinical trial of Moderate Drinking with less dependent drinkers, we did not find a relationship between number of sessions logged in and outcomes [9,35].

On the other hand, Carroll and colleagues did find an additive benefit from their computer-delivered intervention, Computer-Based Training for Cognitive Behavioral Therapy (CBT4CBT) [36]. Their study population, however, was with individuals seeking treatment for substance dependence at a community clinic, which is a population different from individuals seeking assistance online who are not entering treatment for substance abuse.

The prevalence of women (60%) in this study is also consistent with our previous studies of Moderate Drinking (56%) and of our brief motivational intervention, the Drinker's Check-up (48%) [37]. This is remarkable given the epidemiological data indicating that the ratio of problem drinkers by gender is 65% male and 35% female [38], although it does reflect findings that the prevalence of problem drinking among women is increasing [39].

Limitations

There are a number of limitations to this study. First, we did not have a no-intervention control group. While we found it neither practically nor ethically feasible to include such a group in our study, the lack of such a comparison group prevents us from being assured that the treatment assigned was the cause of the improvement. Second, we could not separate out the effects of assessment reactivity that, based on participants' anecdotal reports, did sometimes occur as a function of the baseline evaluation. Third, study participants had, on average, a high level of education (mean 16 years). While this seems to be consistent with the heavy drinkers who affiliate with SMART Recovery, it potentially limits the generalizability of the outcomes in populations with lower levels of education. Fourth, the requirement for an SO to corroborate the participant's self-report of drinking may have further limited the sample. We considered that requirement necessary though as we had no other way to confirm participants' self-reports of their drinking.

Conclusions

Both the Overcoming Addictions Web application and the use of the meetings and other resources of SMART Recovery are effective in helping people recover from heavy problem drinking.

Conflicts of Interest

The senior author holds the copyright and patent pending to the Overcoming Addictions Web application. His plan is to make it available to the general public in September, 2013, on a subscription basis with a portion of the proceeds going to SMART Recovery. He is also a member of SMART Recovery's International Board of Advisors, which is a volunteer position.

Multimedia Appendix 1

Overcoming Addictions overview video.

[MP4 File (MP4 Video), 149MB - jmir_v15i7e134_app1.mp4]

Multimedia Appendix 2

Cost-benefit analysis video, Part 1.

[MP4 File (MP4 Video), 13MB - jmir_v15i7e134_app2.mp4]

Multimedia Appendix 3

Cost-benefit analysis video, Part 2.

[MP4 File (MP4 Video), 8MB - jmir_v15i7e134_app3.mp4]

Multimedia Appendix 4

Cost-benefit analysis video, Part 3.

[MP4 File (MP4 Video), 11MB - jmir_v15i7e134_app4.mp4]

Multimedia Appendix 5

Cost-benefit analysis video, Part 4.

[MP4 File (MP4 Video), 7MB - jmir_v15i7e134_app5.mp4]

Multimedia Appendix 6

The ABC exercise video.

[MP4 File (MP4 Video), 11MB - jmir_v15i7e134_app6.mp4]

Multimedia Appendix 7

CONSORT-EHEALTH checklist V1.6.2 [40].

[PDF File (Adobe PDF File), 1003KB - jmir_v15i7e134_app7.pdf]

References

1. Khadjesari Z, Murray E, Hewitt C, Hartley S, Godfrey C. Can stand-alone computer-based interventions reduce alcohol consumption? A systematic review. Addiction 2011 Feb;106(2):267-282. [doi: 10.1111/j.1360-0443.2010.03214.x] [Medline: 21083832]
2. White A, Kavanagh D, Stallman H, Klein B, Kay-Lambkin F, Proudfoot J, et al. Online alcohol interventions: a systematic review. J Med Internet Res 2010 Dec;12(5):e62 [FREE Full text] [doi: 10.2196/jmir.1479] [Medline: 21169175]
3. Bewick BM, Trusler K, Barkham M, Hill AJ, Cahill J, Mulhern B. The effectiveness of web-based interventions designed to decrease alcohol consumption--a systematic review. Prev Med 2008 Jul;47(1):17-26. [doi: 10.1016/j.ypmed.2008.01.005] [Medline: 18302970]
4. Cunningham JA, Wild TC, Cordingley J, van Mierlo T, Humphreys K. A randomized controlled trial of an internet-based intervention for alcohol abusers. Addiction 2009 Dec;104(12):2023-2032 [FREE Full text] [doi: 10.1111/j.1360-0443.2009.02726.x] [Medline: 19922569]
5. Vernon ML. A review of computer-based alcohol problem services designed for the general public. J Subst Abuse Treat 2010 Apr;38(3):203-211 [FREE Full text] [doi: 10.1016/j.jsat.2009.11.001] [Medline: 20015607]
6. Riper H, Spek V, Boon B, Conijn B, Kramer J, Martin-Abello K, et al. Effectiveness of E-self-help interventions for curbing adult problem drinking: a meta-analysis. J Med Internet Res 2011 Jun;13(2):e42 [FREE Full text] [doi: 10.2196/jmir.1691] [Medline: 21719411]

7.   Cunningham JA. Comparison of two internet-based interventions for problem drinkers: randomized controlled trial. J Med Internet Res 2012;14(4):e107 [FREE Full text] [doi: 10.2196/jmir.2090] [Medline: 22954459]

8.   Postel MG, de Haan HA, ter Huurne ED, Becker ES, de Jong CA. Effectiveness of a web-based intervention for problem drinkers and reasons for dropout: randomized controlled trial. J Med Internet Res 2010 Dec;12(4):e68 [FREE Full text] [doi: 10.2196/jmir.1642] [Medline: 21163776]

9.   Hester RK, Delaney HD, Campbell W. ModerateDrinking.Com and moderation management: outcomes of a randomized clinical trial with non-dependent problem drinkers. J Consult Clin Psychol 2011 Apr;79(2):215-224 [FREE Full text] [doi: 10.1037/a0022487] [Medline: 21319896]

10.   Institute of Medicine. Broadening the base of treatment for alcohol problems. Washington, DC: National Academy Press; 1990.

11.   Horvath AT. SMART Recovery - Self Help for Substance Abuse and Addiction. 2013. URL: http://www.smartrecovery.org/ [accessed 2013-01-28] [WebCite Cache ID 6E14ouBoF]

12.   Horvath AT. SMART Recovery - News. Beliefs about Addiction as a Disease. 2013. URL: http://www.smartrecovery.org/ resources/library/Newsletters/Newsletters/fall2010news&views.pdf [accessed 2013-03-12] [WebCite Cache ID 6F4DTv9Fz]

13.   Hester RK, Reid K. Overcoming Addictions. 2013. URL: http://www.overcomingaddictions.net/ [accessed 2013-01-28] [WebCite Cache ID 6E14VlRgl]

14.   Bowen S, Chawla N, Collins SE, Witkiewitz K, Hsu S, Grow J, et al. Mindfulness-based relapse prevention for substance use disorders: a pilot efficacy trial. Subst Abus 2009;30(4):295-305 [FREE Full text] [doi: 10.1080/08897070903250084] [Medline: 19904665]

15.   Miller WR, Zweben A, DiClemente CC, Rychtarik RG. Motivational Enhancement Therapy Manual: A Clinical Research Guide for Therapists Treating Individuals With Alcohol Abuse and Dependence. Rockville, Maryland: Diane Pub Co; 1995.

16.   Marlatt GA, Baer JS, Donovan DM, Kivlahan DR. Addictive behaviors: etiology and treatment. Annu Rev Psychol 1988;39:223-252. [doi: 10.1146/annurev.ps.39.020188.001255] [Medline: 3278676]

17.   Miller WR, Wilbourne PL, Hettema JE. What works? A summary of alcohol treatment outcome research. In: Hester RK, Miller WR, editors. Handbook of alcoholism treatment approaches: effective alternatives. Boston: Allyn and Bacon; 2003:13-63.

18.   Koob GF. Theoretical frameworks and mechanistic aspects of alcohol addiction: alcohol addiction as a reward deficit disorder. Curr Top Behav Neurosci 2013;13:3-30. [doi: 10.1007/7854_2011_129] [Medline: 21744309]

19.   SMART Recovery. URL: http://www.smartrecovery.org/meetings_db/view/ [accessed 2013-03-14] [WebCite Cache ID 6F7Kl74p5]

20.   Connors GJ, Donovan DM, DiClemente CC. Substance abuse treatment and the stages of change: selecting and planning interventions. New York: Guilford Press; 2001.

21.   Meyers RJ, Smith JE. Clinical guide to alcohol treatment: The community reinforcement approach. New York: Guilford Press; 1995.

22.   Ellis A, Velten E. When AA doesn't work for you: rational steps to quitting alcohol. Fort Lee, N.J: Barricade Books; 1992.

23.   Steinberger H. Smart Recovery Handbook. Mentor, OH: Smart Recovery; 2004.

24.   Marlatt GA, Gordon JR. Relapse prevention: maintenance strategies in the treatment of addictive behaviors. New York: Guilford Press; 1985.

25.   Babor T, Higgins-Biddle J, Sauders J, Monteiro M. WHO. 2013. The Alcohol Use Disorders Identification Test URL: http:/ /whqlibdoc.who.int/hq/2001/who_msd_msb_01.6a.pdf [accessed 2013-01-30] [WebCite Cache ID 6E43o6Bg6]

26.   Cherpitel CJ. Screening for alcohol problems in the U.S. general population: comparison of the CAGE, RAPS4, and RAPS4-QF by gender, ethnicity, and service utilization. Rapid Alcohol Problems Screen. Alcohol Clin Exp Res 2002 Nov;26(11):1686-1691. [doi: 10.1097/01.ALC.0000036300.26619.78] [Medline: 12436057]

27.   Derogatis LR. The Brief Symptom Inventory. Minneapolis, MN: Pearson Assessments; 2000.

28.   Association for Behavioral and Cognitive Therapies. ABCT Find A Therapist Service. 2013. URL: http://www.abct.org/ Members/?m=FindTherapist&fa=FT_Form&nolm=1 [accessed 2013-01-28] [WebCite Cache ID 6E1595dfP]

29.   Sobell LC, Maisto SA, Sobell MB, Cooper AM. Reliability of alcohol abusers' self-reports of drinking behavior. Behav Res Ther 1979;17(2):157-160. [Medline: 426744]

30.   Sobell L, Sobell M. Alcohol Timeline Followback User's Manual. In: Timeline followback: user's guide. Toronto: Addiction Research Foundation = Fondation de la recherche sur la toxicomanie; 1996.

31.   Miller WR, Tonigan JS, Longabaugh R. The Drinker's Inventory of Consequences (DrInC): An instrument for assessing adverse consequences of alcohol abuse. Rockville, Maryland: National Institute on Alcohol Abuse and Alcoholism; 1995.

32.   Little RJA. Missing-data adjustments in large surveys. Journal of Business & Economic Statistics 1988;6(3):287-296.

33.   Hester RK. The Drinker's Check-up. 2013. URL: http://www.drinkerscheckup.com [accessed 2013-01-29] [WebCite Cache ID 6E2g8SqJB]

34.   Hester RK, Squires DD. Web-based norms for the Drinker Inventory of Consequences from the Drinker's Checkup. J Subst Abuse Treat 2008 Oct;35(3):322-327. [doi: 10.1016/j.jsat.2007.10.005] [Medline: 18248943]

35.   Hester RK. Moderate Drinking. 2013. URL: http://www.moderatedrinking.com/home/default_home.aspx?p=register_login
       [accessed 2013-01-29] [WebCite Cache ID 6E2gJc5XR]
36.   Carroll KM, Ball SA, Martino S, Nich C, Babuscio TA, Nuro KF, et al. Computer-assisted delivery of cognitive-behavioral
       therapy for addiction: a randomized trial of CBT4CBT. Am J Psychiatry 2008 Jul;165(7):881-888 [FREE Full text] [doi:
       10.1176/appi.ajp.2008.07111835] [Medline: 18450927]
37.   Hester RK, Squires DD, Delaney HD. The Drinker's Check-up: 12-month outcomes of a controlled clinical trial of a
       stand-alone software program for problem drinkers. J Subst Abuse Treat 2005 Mar;28(2):159-169. [doi:
       10.1016/j.jsat.2004.12.002] [Medline: 15780546]
38.   Hasin DS, Stinson FS, Ogburn E, Grant BF. Prevalence, correlates, disability, and comorbidity of DSM-IV alcohol abuse
       and dependence in the United States: results from the National Epidemiologic Survey on Alcohol and Related Conditions.
       Arch Gen Psychiatry 2007 Jul;64(7):830-842. [doi: 10.1001/archpsyc.64.7.830] [Medline: 17606817]
39.   Grucza RA, Bucholz KK, Rice JP, Bierut LJ. Secular trends in the lifetime prevalence of alcohol dependence in the United
       States: a re-evaluation. Alcohol Clin Exp Res 2008 May;32(5):763-770 [FREE Full text] [doi:
       10.1111/j.1530-0277.2008.00635.x] [Medline: 18336633]
40.   Eysenbach G, CONSORT-EHEALTH Group. 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]

Abbreviations

ANCOVA: analysis of covariance
AUDIT: Alcohol User Disorders Identification Test
BSI: Brief Symptom Inventory
DDD: mean standard drinks per drinking day
DrInC: Drinker's Inventory of Consequences
DWI: driving while intoxicated
InDUC: Inventory of Drug Use Consequences
OA: Overcoming Addictions Web application
PDA: percent days abstinent (in previous 90 days)
SMART: Self-Management And Recovery Training
SO: significant other
SR: SMART Recovery meetings
TLFB: Timeline Follow-Back

XSL•FO
RenderX

Original Paper

Supervised Patient Self-Testing of Warfarin Therapy Using an Online System

Luke Ryan Elliot Bereznicki[1*], BPharm, PhD; Shane Leigh Jackson[1*], BPharm, PhD; Gregory Mark Peterson[1*], BPharm, PhD, MBA

Unit for Medication Outcomes Research and Education, School of Pharmacy, University of Tasmania, Hobart, Australia
[*]all authors contributed equally

Corresponding Author:
Luke Ryan Elliot Bereznicki, BPharm, PhD
Unit for Medication Outcomes Research and Education
School of Pharmacy
University of Tasmania
Private Bag 26
Hobart, 7001
Australia
Phone: 61 036226 ext 2195
Fax: 61 0362262870
Email: Luke.Bereznicki@utas.edu.au

Abstract

Background: Point-of-care international normalized ratio (INR) monitoring devices simplify warfarin management by allowing selected patients to monitor their own therapy in their homes. Patient self-testing (PST) has been shown to improve the clinical outcomes of warfarin therapy compared to usual care.

Objective: To compare management of warfarin therapy using PST combined with online supervision by physicians via a custom system with usual warfarin management, which involved laboratory testing and physician dosing.

Methods: Interested patients were recruited via community pharmacies to participate in a warfarin PST training program. Participants were required to have a long-term indication for warfarin, have been taking warfarin for at least 6 months, and have Internet access in their home. The training involved two sessions covering theoretical aspects of warfarin therapy, use of the CoaguChek XS, and the study website. Following training, patients monitored their INR once weekly for up to 3 months. Patients and physicians utilized a secure website to communicate INR values, dosage recommendations, and clinical incidents. Physicians provided a 6-12 month history of INR results for comparison with study results. The percentage of time spent within the therapeutic INR range (TTR) was the primary outcome, with participants acting as their own historical controls. The percentage of INR tests in range and participant satisfaction were secondary outcomes.

Results: Sixteen patients completed training requirements. The mean age of participants was 69.8 (SD 10.1) years. TTR improved significantly from 66.4% to 78.4% during PST (*P*=.01), and the number of tests within the target range also improved significantly (from 66.0% at prior to the study to 75.9% during PST; *P*=.04). Patients and physicians expressed a high degree of satisfaction with the monitoring strategy and online system.

Conclusions: PST supported by an online system for supervision was associated with improved INR control compared to usual care in a small group of elderly patients. Further research is warranted to investigate the clinical outcomes and cost-effectiveness of online systems to support patients monitoring medications and chronic conditions in the home.

XSL•FO

RenderX

Introduction

Warfarin is an anticoagulant that has been, and continues to be, the standard of care to prevent and treat thromboemboli. It is estimated that between 1% and 2% of the population of the developed world currently receives oral anticoagulants, predominantly warfarin, on a regular basis, mainly for the prevention of ischemic stroke associated with chronic atrial fibrillation. While newer anticoagulants have become available for some indications, their high cost, limited range of indications, and uncertain risk/benefit profile, particularly in the elderly, ensure that debate continues regarding their place in therapy as either replacements or alternatives to warfarin [1-3]. While effective, warfarin is a well-known cause of adverse drug events, and it is fundamental that efforts are made to improve the safety of warfarin therapy and maximize its clinical benefits.

Despite decades of clinical use, there are a number of aspects of warfarin therapy that can be targeted to improve patient outcomes. Initial dosing can be optimized through the application of dosing algorithms based on clinical and genetic parameters [4], and patient knowledge of warfarin (often described as suboptimal in the literature) can be improved through health professional intervention [5,6]. Dietary vitamin K intake can be optimized in some patients to improve control [7], and the use of interacting medications can be minimized through judicious prescribing. While the need for regular monitoring of the international normalized ratio (INR), necessitated by large interindividual differences in response to warfarin, can be a burden to patients and health care systems, the ready availability of INR testing means that warfarin therapy can be rigorously monitored to optimize patient safety. The degree of INR control is the major determinant of the efficacy and safety of warfarin therapy [8], and optimized monitoring may make warfarin more clinically effective and cost-effective than its competitors [9,10]. Of course, if warfarin therapy is not well controlled, patients are at increased risk of thrombotic and hemorrhagic complications [11].

Traditionally, warfarin is managed by primary care physicians, pathology providers, or by health professionals in dedicated anticoagulation clinics. Point-of-care (POC) devices offer an accurate alternative to laboratory monitoring [12,13], and their availability has led to an increased focus on patient self-testing (PST) of warfarin. This may comprise self-testing, where the result of the test is communicated to a physician for management (usually by phone), or self-management, where the patient self-adjusts the warfarin dose. PST has been evaluated in a number of well-controlled studies, and a recent meta-analysis demonstrated that self-monitoring of warfarin therapy results in significant reductions in the incidence of thromboembolic complications [14]. The reasons that PST can be more effective than usual care are multifactorial and include enabling an increase in testing frequency, educating patients about important aspects of warfarin therapy, and empowering patients to take a greater role in their own care.

It is possible that the appropriate application of information technology could greatly improve warfarin management by improving communication between patients and physicians and facilitating self-monitoring. The Internet offers significant promise as an enabling factor for PST by allowing patients to monitor their therapy at home and receive ongoing advice without having to visit their supervising health professional, particularly in situations where patients and/or their supervisors may be uncomfortable with the patient taking unsupervised control of their condition.

We sought to develop and test an approach to improve INR control by training patients to self-test their INR and linking them to their primary care physician with an online support system. Our hypothesis was that self-testing with online support would improve INR control compared to usual care. This required the development of a training process to enable participants to self-test and an online system to link patients with their physicians and support staff. We conducted a prospective, proof-of-concept study to compare the INR control achieved with online-supervised PST to usual care, investigate patient and physician views on the online model of care, and provide the foundation for more extensive research in this area.

Methods

Design

This was a prospective study of PST with online decision support. Participants acted as their own historical controls. The INR control achieved with PST was compared with the INR control achieved in the 6-12 months immediately preceding the study using conventional management (laboratory INR with physician dose adjustment).

Patient Recruitment

To participate in this study, patients had to have been taking warfarin for at least 6 months, have a long-term indication for warfarin therapy, be willing to participate in training to enable PST, and have Internet access in the home. To gauge patient interest in PST, we randomly selected 20 community pharmacies in southern Tasmania to participate in a preliminary survey (using random number generation and a list of pharmacies). These pharmacies contacted all patients who had been dispensed warfarin in the past 3 months and mailed them an information sheet, which invited patients to contact the researchers if they were interested in participating in a survey regarding their interest in PST. A follow-up survey of patients who indicated a willingness to be contacted to provide additional information was conducted to identify patients who met the inclusion criteria. These patients were approached to provide their consent to be involved in the study. Once consent was provided, the research team contacted the patient's primary care physician informing them of the study and their patient's interest, and seeking their consent to participate. Physician consent was essential for the patient to be included in the trial. Ethics approval was provided by the Southern Tasmania Health and Medical Human Research Ethics Committee.

Patient Training

Once enrolled in the trial, patients were trained to use a portable INR monitor (CoaguChek XS, Roche Diagnostics Australia) by the research team. The education consisted of two sessions

of 1-2 hours duration held approximately 1 week apart. The first session covered background information on warfarin, risk of bleeding, diet, the INR, and a practical demonstration of the CoaguChek XS, and was conducted as a small group session (2-6 patients). Patients then received their own monitor and were asked to conduct 4-5 tests prior to the next session to become accustomed to the device and confident in their testing technique.

A "run-in" phase, where patients completed 2 INR tests on the CoaguChek XS in conjunction with 2 pathology tests to compare for accuracy, ensured that the research team, the physician, and the patient were satisfied that the monitor provided accurate results and could be used effectively. Patients were asked to complete these comparison tests prior to the second education session. If the CoaguChek INR results were not suitably accurate (defined as the CoaguChek XS INR being within 15% of the laboratory INR) on both occasions, further instruction was provided. If further comparisons were not suitably accurate, the patient was excluded from the trial. The second session covered other aspects of using the monitor, such as quality control and using the online system to relay the results to the physician. This session was conducted as a home visit by one of the researchers to ensure effective use of the patient's home computer. The physicians also received some instruction on how to access and utilize the online system. This education was delivered at the physician's surgery in a one-on-one session with a member of the research team.

An observed assessment was made of each participant before home monitoring could occur. Patients were required to demonstrate that they could use the CoaguChek XS in a proficient manner and satisfactorily complete a test based on the theory content of the course.

Website Development and Functionality

A consultant information technology company, in collaboration with the research team, developed the online system used in the study. In order to tailor the website to the needs of the end users, the research team conducted two focus group meetings, with patients and physicians. The facilitated discussions were aimed at improving the design of the website and identifying, in particular, which functions were required, flow of information, and training requirements.

The self-testing and data entry procedure consisted of performance of the INR test on the POC INR monitor, logging on to the secure website, and manual entry of the test result. The system provided an overview of the steps required in the testing process and asked users to indicate if they had conducted the test appropriately. Following entry of the result by the patient, the system displayed the result and asked the patient to confirm its accuracy. The system screened the result and noted whether the test was below, within, or above the prescribed INR range and asked the patient to document any changes in diet, medication, or make any additional comments. Finally, the system requested confirmation of all information entered and sent an alert to the physician via email.

Physicians were asked to respond to the test result within 24 hours and provide the warfarin dose and date of the next test.

Physicians were free to log on to the secure website and view patient details at any stage. Physicians were able to make a recommendation based on the INR, information provided by the patient, and stored history. The system notified the patient of the dosage recommendation when it was available from the physician via email. The system was used by physicians to set the date of the next test and alerted support staff if tests became overdue. If a patient had not completed an INR test within 24-48 hours of the test being due, the researchers contacted the patient. Similarly, the physician was to be contacted if the patient had not received a dosage recommendation within 24-48 hours of completing the INR test. Telephone support was available from the research team at all times.

The online system allowed patients to access to a variety of educational materials related to anticoagulation, and dietary and lifestyle advice for patients taking warfarin. Patients were also able to visualize their INR results on a graph and view physician recommendations. The system stored patient information, including contact details and INR history; access to this information by patients and physicians was only available through a password-protected website. Patients had access to only their own details while physicians were able to access information relevant to all patients under their care. Support staff had access to all information. Communications to and from this site were encrypted. Emails did not contain any sensitive information. All data were physically stored in a secure environment and treated confidentially and anonymously. For screenshots of the online system, see Multimedia Appendix 1.

International Normalized Ratio Testing

Once training was complete, participants tested their INR approximately once per week, or more or less frequently if required by the physician, for a period of up to 3 months. The percentage of time in therapeutic range (TTR) and percentage of tests within the INR target range were determined for each patient during the trial. This was compared to their previous level of control for the 6-12 months immediately prior to the commencement of the PST phase of the study (provided at study entry with each participant's consent). The function to calculate the TTR was based on the method of linear interpolation originally proposed by Rosendaal et al [15].

Sample Size and Statistical Analysis

The primary outcome was change in TTR from baseline. The TTR for each participant was determined for prePST and PST data. Paired t tests were used to determine if any significant change had occurred compared to baseline. Statistical significance was set at $P<.05$. Feedback from physicians and patients was sought regarding the system and its ease of use. This was obtained using evaluation questionnaires featuring visual analogue scales, ranging from 0 to 10. Scores <5 indicated disagreement with the statements provided, while scores >5 indicated agreement with the statements provided. The study was not powered to detect statistical differences in clinical outcomes, such as bleeds, although their occurrence was documented as a matter of course. All information was stored and analyzed using SPSS 19.0 (version 19.0 for Mac).

A sample size of approximately 20 patients was deemed be adequate to demonstrate the feasibility of this type of warfarin management. The literature suggests that patients in the community spend 50-60% of their time within the target range [16]. It was envisaged that this could be improved to 75% with the intervention based on a prior study involving a similar intervention [17]. At a power of 80% and statistical significance set at 0.05, 16 patients analyzed before and after were required.

Results

Participants

Figure 1 shows a summary of the recruitment process. Of the 832 patients contacted by their community pharmacies, 243 (29.2%) contacted the researchers. One hundred and sixty-eight patients returned the survey (69.1%), of whom 122 (72.6%) indicated a willingness to undertake training to enable PST. A follow-up survey of 66 of these patients who indicated a willingness to be contacted to provide additional information identified 28 potential participants in the study who reported having Internet access in their home and met the other inclusion criteria. Twenty-two of the 28 patients (78.6%) who met the inclusion criteria consented to participate in the trial. Five patients did not complete the training requirements due to logistical issues, and 1 patient did complete some of the training requirements but was unable to continue in the study due to the unavailability of their physician. Sixteen patients completed the initial training program and went on to perform PST. Sixteen different physicians were involved in the management of these patients. Once training was complete, there were no withdrawals from the study. Table 1 displays patient characteristics for the patients who performed PST.

Device Accuracy

A total of 59 comparison INRs (CoaguChek XS and laboratory INR conducted within 4 hours of each other) were completed either on entry into, or during the trial by the participants. The CoaguChek XS INR values were significantly correlated with the laboratory INR values (r=0.91, P=.01). The mean difference in INR (laboratory minus CoaguChek XS) was 0.07 (SD 0.06) (t_{58}=2.56, P=.01).

International Normalized Ratio Control

The mean TTR prior to PST was 66.4% (SD 17.7%). A total of 309 INR tests were provided with an average of 19.3 (SD 7.9) tests per patient. The mean duration of time encompassed by the baseline data was 338.4 (SD 52.8) days. During PST, patients had a mean TTR of 78.4% (SD 20.1%). The mean number of results per patient was 7.5 (SD 3.0; 120 home tests were completed by the cohort). The mean duration of PST was 45.1 (SD 16.0) days. Figure 2 shows a comparison of mean TTR during usual care and PST. There was a statistically significant improvement in the TTR when patients performed PST (mean improvement 12.0% (SD 17.3%), P=.01) (Table 2). The percentage of time spent below and above target range were reduced nonsignificantly. Thirteen of 16 (81.3%) patients improved on their baseline control (Figure 3). The mean increase in the number of tests in target range was 9.9% (66.0%, SD 16.6% usual care, 75.9%, SD 19.2% during intervention, P=.04), a significant improvement. No clinical outcomes (events of major bleeding or thromboembolism) were observed during the intervention.

Participant Evaluation

At the completion of the trial, questionnaires were sent to all physicians and patients. The response rate for physicians was 87.5% (14/16) and for patients was 93.8% (15/16).

Table 1. Patient characteristics.

Characteristic		n (%)
Male sex		12 (75.0)
Mean age (SD years)		69.8 (10.1)
Indication for warfarin		
	Atrial fibrillation	7 (43.8)
	Venous thrombosis	2 (12.5)
	Heart valve	6 (37.5)
	Other	1 (6.3)
Target INR range		
	2.0-3.0	10 (62.5)
	2.5-3.5	2 (12.5)
	Other (specified by physician)	4 (25.0)

Figure 1. Patient recruitment.
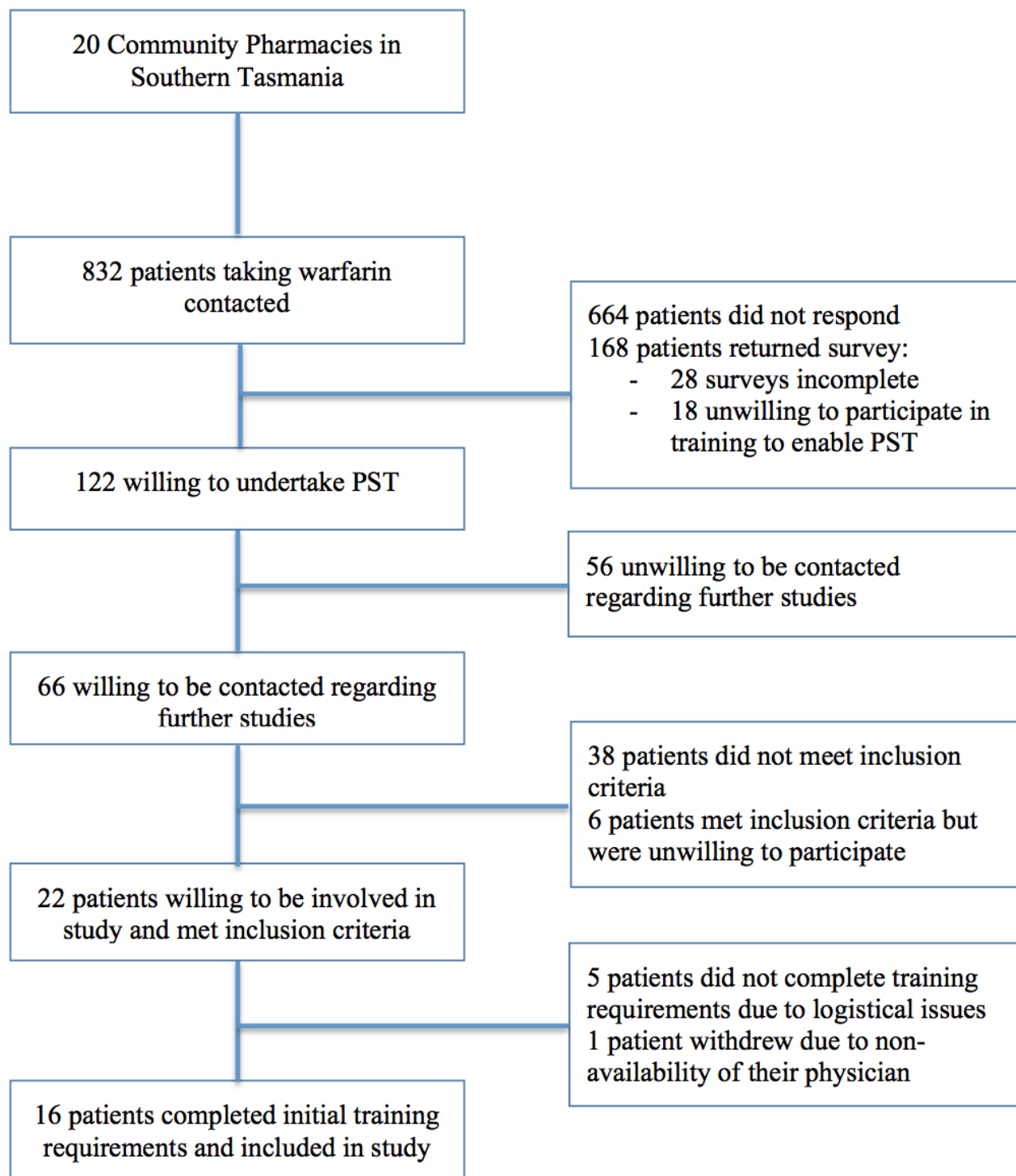
Table 2. Percentage of time in target range.

	Usual care	Patient self-testing	Mean change	P
% within range (SD)	66.4 (17.7)	78.4 (20.1)	+12.0 (17.3)	.01
% below range (SD)	22.2 (24.3)	13.5 (15.5)	-8.7 (20.2)	.11
% above range (SD)	11.4 (10.2)	8.1 (11.1)	-3.2 (14.0)	.37

Physician Evaluation

Physicians indicated that they found the intervention to be a beneficial service for their patients (median score 7.5, range 6.2-8.9). They were also positive when asked whether they would feel more confident in managing patients taking warfarin if it was a regular service (median score 7.0, range 3.0-8.3). Physicians strongly agreed that they were confident in the accuracy of POC INR results (median score 7.5, range 5.0-9.3) and tended to agree that the system was easy to use (median score 7.0, range 4.7-9.5). Physicians generally found that receiving, reviewing, and responding to an INR result took 1-3 minutes. Most physicians responded positively (median score of 8.0, range 2.6-9.5) when asked if they saw this model of care as a feasible way to manage patients taking warfarin in the future, and believed that more patients would benefit from this service (median score 7.5, range 5.0-8.4).

Patient Evaluation

All patients who responded to the evaluation questionnaire found the intervention to be a worthwhile service (median score 9.5, range 5.5-9.5) and would feel more confident about taking warfarin if it was offered as a regular service (median score 7.0, range 5.5-9.5). Most patients indicated they preferred self-testing compared to laboratory INR testing (median score 7.5, range 4.5-9.5). Patients felt strongly that the education and training provided was of benefit to them (median score 9.5, range 6.9-9.5) and generally felt that their participation had improved their knowledge regarding their treatment (median score 7.5, range 5.2-9.5). They felt that the portable monitor was easy to use (median score 7.5, range 5.5-9.5) and were confident in its accuracy (median score 9.3, range 5.7-9.5). They also reported that the website was easy to use (median score 7.5, range 5.7-9.5) and were satisfied with the support provided (median score 9.5, range 7.5-9.5). They reported that they generally spent between 1 and 3 minutes entering INR results on the system or checking the dose changes provided by their physicians.

Figure 2. Individual percentage of time in therapeutic range (TTR) during patient self-testing (PST) and usual care (UC) (the dotted line shows the mean TTR for each management approach).
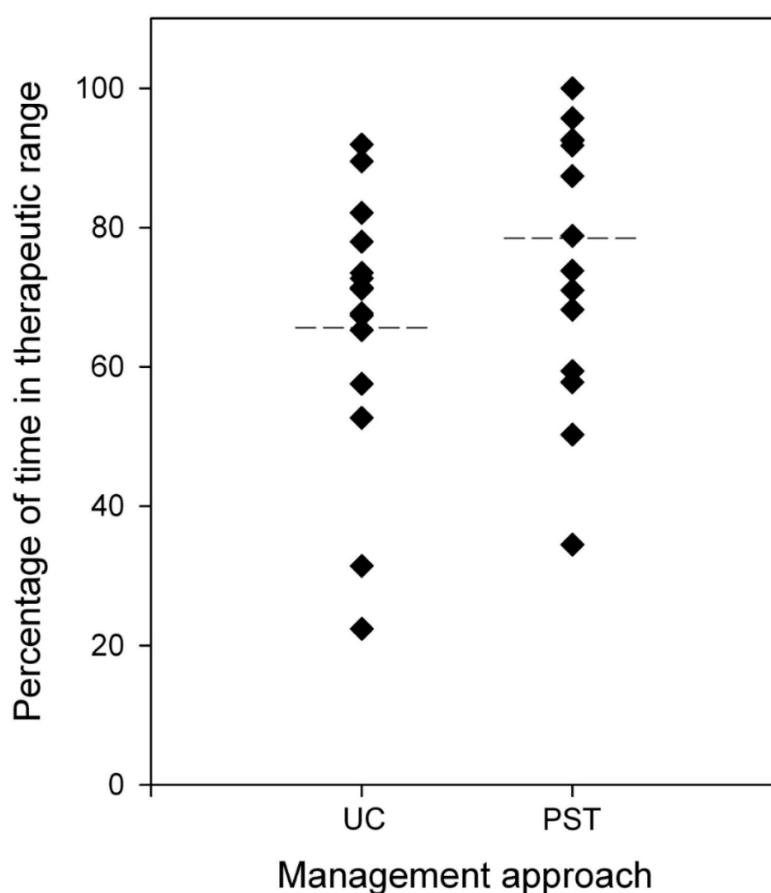
Figure 3. Percentage of time in therapeutic range for patient self-testing vs usual care.
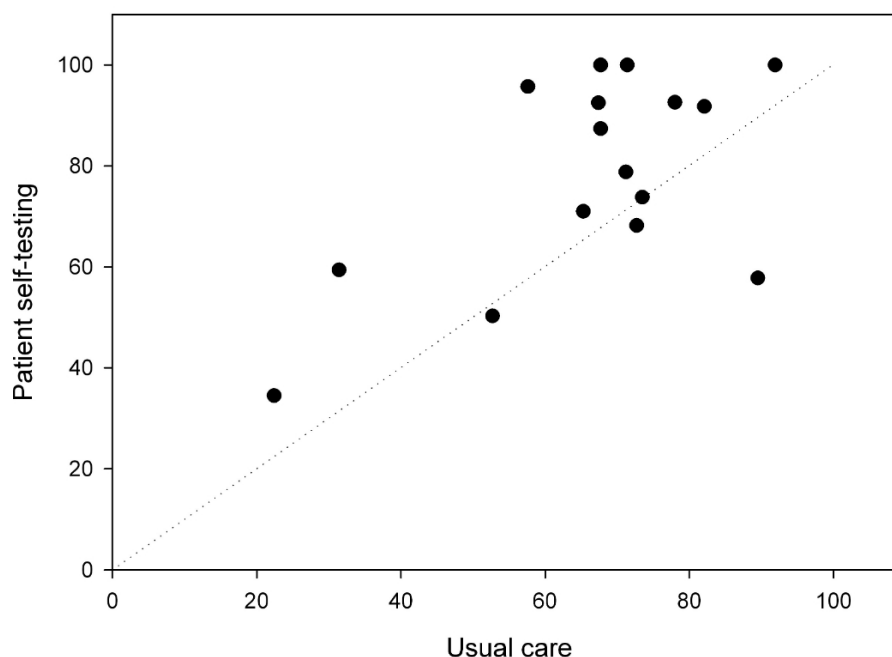
Discussion

International Normalized Ratio Control

This study found that INR control improved when patients performed PST and were remotely supervised by their physicians using a custom online system, compared to their usual care under the same physician in a small number of patients. Participants received training and used the CoaguChek XS to monitor their INR, entered their INR into an online system, and received advice from their supervising physician. Patients and physicians alike found it to be a valuable model of care for warfarin therapy.

There is a direct relationship between TTR and clinical outcomes for patients taking warfarin [11,18]. The generalizability of the results of international trials comparing new anticoagulants to warfarin depends largely on the TTR achieved in the trials and the TTR achieved by patients taking warfarin within the particular health care systems studied. It has been established that the relative efficacy and safety of comparator drugs to warfarin varies depending on the quality of INR control [9]. Therefore, the TTR is critical in determining the relative efficacy, safety, and potential cost-effectiveness of new anticoagulants compared to warfarin, as well as comparing various models of warfarin management. We found that TTR improved from 66% to 78% with the model of care developed in this study. The usual care mean TTR of 66% was higher than we expected; a mean TTR in the order of 50-60% was anticipated based on previous community-based nonrandomized studies [16]. Interestingly, the usual care mean TTR was higher than that achieved in several recent randomized trials comparing warfarin to new anticoagulants, even though the TTR in the trials is often said to be higher than that achieved in practice [19-21]. This is possibly due to the recruitment process, which identified patients who were interested in PST and were perhaps more motivated and knowledgeable regarding their treatment

than other people taking warfarin. However, it is notable that even in a group with relatively good baseline INR control, it was possible to achieve a TTR approaching 80% with PST. Improvements in TTR of this magnitude are likely to be associated with improved clinical outcomes. A retrospective study in patients with atrial fibrillation found that a 7% improvement in TTR is associated with 1 fewer hemorrhagic event per 100 patient-years and a 12% improvement is associated with 1 fewer thromboembolic event per 100 patient-years [22].

Self-monitoring often results in improved INR control compared to the control achieved in comparator groups (either primary care or anticoagulation clinic management) [14]. A number of studies have shown improved TTR [23-25], while in other studies PST has not resulted in an increase in TTR but measures of stability have improved [26,27]. TTR may not be the only means by which PST might result in improved clinical outcomes. In a trial by Menedez et al [26], TTR did not improve significantly between the usual care group and the self-managing group but the incidence of major warfarin-related complications was reduced significantly. The beneficial effects of PST may therefore relate to patient education [6], patient empowerment [28,29], or improved medication adherence [30], in addition to improving measures of INR control.

The INR testing frequency in this study was weekly unless otherwise specified by the supervising physician. In the prePST phase, the mean testing frequency was approximately 18 days. Some researchers argue that the improvements in TTR generally associated with PST are largely due to an increased testing frequency [31-33]. This may not necessarily be the case as PST has been shown to improve the TTR without a change in testing frequency [34], and another study recently found that a longer testing interval (12 weeks) was not associated with any change in mean TTR compared to a 4-week testing interval [35]. In fact, more frequent testing may actually have a detrimental effect on TTR, as it may lead to unnecessary dose adjustment

[35]. Regardless, our study was not designed to differentiate between the potential individual benefits of patient education, increased testing frequency, PST, or the presence of the online system.

PST provides the flexibility to monitor more or less frequently at the discretion of the physician without creating undue pressure on the patient to attend pathology testing and/or physician consultations. A weekly testing frequency can also be achieved with PST at a similar cost to monthly laboratory testing. In a Canadian study, the ongoing costs associated with weekly self-testing were identical to the costs of 4-weekly conventional monitoring [36].

Comparison With Prior Work

Our results are consistent with the improvements in TTR associated with PST and online systems in three other studies [17,37,38]. The studies by O'Shea et al [17] and Ryan et al [37] used the same Internet-based system, which provided a decision-support tool to assist patients adjust their own warfarin dose. The improvements in INR control achieved in each of these studies were similar. The TTR improved from 63% to 74% ($P<.01$) in the pilot investigation [17] and from 59% to 74% ($P<.01$) in the subsequent randomized controlled trial [37]. The study by Harper and Pollock used a different online system to support PST [38]. This system provided instant feedback on the warfarin dose if the INR was in the therapeutic range, but if outside the range sent an alert to the physician to review. The TTR improved nonsignificantly from 72% to 81%. Unlike in these studies, the system used in our study was not designed to provide any dosing advice to patients—in all instances, the test result was sent to physicians for review. It is interesting that this difference did not appear to affect the TTR achieved with PST and the online system, although a larger comparative study would be required to verify this observation.

Our study is noteworthy for two other reasons. First, our patients were older than those studied in previous studies involving online systems. The mean age of our participants was 70 years, while the median age was 54 years in the study by O'Shea [17] and the mean age was 59 years in the larger study by Ryan [37]. Our patients were also older than those in almost all previous studies of PST [14]. This is important because warfarin is increasingly indicated in an older population, but advanced age is often seen as a deterrent to warfarin therapy because of a perceived increase in the risk of bleeding [39]. Our results suggest that not only can excellent INR control be achieved in older patients, they can also be trained to successfully perform PST and use online systems. Our participants were a selected population; nonetheless, the results indicate that advanced age should not necessarily be considered a deterrent to achieving excellent INR control, utilizing PST or online systems. Clearly, these observations need to be tested in larger, long-term studies.

Second, we have provided data on the experience of our end users, both physicians and patients. Our online system was designed with input from end users—this led to a focus on ease of use, convenience, and safety based on their priorities. The physicians who were involved in the pilot study and completed an evaluation questionnaire all found it to be a valuable service for their patients. Physicians agreed that more patients would benefit from this type of service and the percentage of patients that they felt this system would be suitable for ranged from 12% to 98%. Our experience in this study suggests that most of these patients would be capable of completing the necessary training to self-test and use the online system. Previous research in the United Kingdom using an unselected population gave similar results in terms of capability to self-test [40]. As far as using the system was concerned, physicians responded that the system was easy to use and the warfarin home monitoring website was easy to navigate. The majority of physicians did not require any additional support following their initial training.

Importantly, the patients using the online system found it to be a valuable service that made them feel more confident about their warfarin therapy. They found the initial training beneficial and also agreed that their warfarin knowledge had improved as a result of the training. Importantly for the ongoing development of the system, patients found the website training easy and were highly satisfied with the ongoing support and by their physicians' involvement and use of the system. Patients reported spending the same amount of time on the warfarin home monitoring trial website as their physicians, that is, 1-3 minutes per test. When asked whether they would prefer to monitor their warfarin therapy at home, patients indicated that, in general, they would.

Limitations

Our study involved a selected group of participants who may not be representative of the broader population of people taking warfarin. It is possible that they were more motivated and possibly more adherent with their therapy than other patients [33]. Other potential limitations of the study include the small sample size, nonrandomized design, and relatively short duration of the intervention.

Conclusions

This proof-of-concept study was successful in demonstrating the feasibility of an alternative warfarin management strategy involving supervised PST using an online system in a small group of selected participants. Patients spent more than 78% of time in the therapeutic range while self-testing, which was a significant improvement from their previous INR control. Patients and physicians were highly satisfied with the monitoring system. Further research is warranted to investigate the benefits and implications of this strategy for people taking warfarin, as well as other narrow-therapeutic index drugs and those with chronic diseases where regular monitoring is indicated.

XSL•FO

RenderX

the online system used in this project. This project was supported by the Tasmanian Electronic Commerce Centre under the Business Development Fund, an element of the Launceston Broadband Project, funded by the Commonwealth through the Department of Communications, Information Technology and the Arts.

Conflicts of Interest

Multimedia Appendix 1

Screenshots of the Web-based system.

[PDF File (Adobe PDF File), 542KB - jmir_v15i7e138_app1.pdf]

References

1. Beyth RJ, Landefeld CS. Learning the respective roles of warfarin and dabigatran to prevent stroke in patients with nonvalvular atrial fibrillation. Ann Intern Med 2011 Nov 15;155(10):714-5, W218. [doi: 10.1059/0003-4819-155-10-201111150-00012] [Medline: 22084336]
2. Legrand M, Mateo J, Aribaud A, Ginisty S, Eftekhari P, Huy PT, et al. The use of dabigatran in elderly patients. Arch Intern Med 2011 Jul 25;171(14):1285-1286. [doi: 10.1001/archinternmed.2011.314] [Medline: 21788545]
3. Ansell J. New oral anticoagulants should not be used as first-line agents to prevent thromboembolism in patients with atrial fibrillation. Circulation 2012 Jan 3;125(1):165-70; discussion 170 [FREE Full text] [doi: 10.1161/CIRCULATIONAHA.111.031153] [Medline: 22215891]
4. Epstein RS, Moyer TP, Aubert RE, O Kane DJ, Xia F, Verbrugge RR, et al. Warfarin genotyping reduces hospitalization rates results from the MM-WES (Medco-Mayo Warfarin Effectiveness study). J Am Coll Cardiol 2010 Jun 22;55(25):2804-2812. [doi: 10.1016/j.jacc.2010.03.009] [Medline: 20381283]
5. Tang EO, Lai CS, Lee KK, Wong RS, Cheng G, Chan TY. Relationship between patients' warfarin knowledge and anticoagulation control. Ann Pharmacother 2003 Jan;37(1):34-39. [Medline: 12503930]
6. Khan TI, Kamali F, Kesteven P, Avery P, Wynne H. The value of education and self-monitoring in the management of warfarin therapy in older patients with unstable control of anticoagulation. Br J Haematol 2004 Aug;126(4):557-564. [doi: 10.1111/j.1365-2141.2004.05074.x] [Medline: 15287950]
7. Rombouts EK, Rosendaal FR, van der Meer FJ. Influence of dietary vitamin K intake on subtherapeutic oral anticoagulant therapy. Br J Haematol 2010 May;149(4):598-605. [doi: 10.1111/j.1365-2141.2010.08108.x] [Medline: 20151978]
8. Odén A, Fahlén M. Oral anticoagulation and risk of death: a medical record linkage study. BMJ 2002 Nov 9;325(7372):1073-1075 [FREE Full text] [Medline: 12424167]
9. Wallentin L, Yusuf S, Ezekowitz MD, Alings M, Flather M, Franzosi MG, RE-LY investigators. Efficacy and safety of dabigatran compared with warfarin at different levels of international normalised ratio control for stroke prevention in atrial fibrillation: an analysis of the RE-LY trial. Lancet 2010 Sep 18;376(9745):975-983. [doi: 10.1016/S0140-6736(10)61194-4] [Medline: 20801496]
10. Shah SV, Gage BF. Cost-effectiveness of dabigatran for stroke prophylaxis in atrial fibrillation. Circulation 2011 Jun 7;123(22):2562-2570 [FREE Full text] [doi: 10.1161/CIRCULATIONAHA.110.985655] [Medline: 21606397]
11. White HD, Gruber M, Feyzi J, Kaatz S, Tse HF, Husted S, et al. Comparison of outcomes among patients randomized to warfarin therapy according to anticoagulant control: results from SPORTIF III and V. Arch Intern Med 2007 Feb 12;167(3):239-245. [doi: 10.1001/archinte.167.3.239] [Medline: 17296878]
12. Kaatz SS, White RH, Hill J, Mascha E, Humphries JE, Becker DM. Accuracy of laboratory and portable monitor international normalized ratio determinations. Comparison with a criterion standard. Arch Intern Med 1995 Sep 25;155(17):1861-1867. [Medline: 7677552]
13. Jackson SL, Bereznicki LR, Peterson GM, Marsden KA, Jupe DM, Tegg E, et al. Accuracy, reproducibility and clinical utility of the CoaguChek S portable international normalized ratio monitor in an outpatient anticoagulation clinic. Clin Lab Haematol 2004 Feb;26(1):49-55. [Medline: 14738438]
14. Heneghan C, Ward A, Perera R, Self-Monitoring Trialist Collaboration, Bankhead C, Fuller A, et al. Self-monitoring of oral anticoagulation: systematic review and meta-analysis of individual patient data. Lancet 2012 Jan 28;379(9813):322-334. [doi: 10.1016/S0140-6736(11)61294-4] [Medline: 22137798]
15. Rosendaal FR, Cannegieter SC, van der Meer FJ, Briët E. A method to determine the optimal intensity of oral anticoagulant therapy. Thromb Haemost 1993 Mar 1;69(3):236-239. [Medline: 8470047]
16. van Walraven C, Jennings A, Oake N, Fergusson D, Forster AJ. Effect of study setting on anticoagulation control: a systematic review and metaregression. Chest 2006 May;129(5):1155-1166. [doi: 10.1378/chest.129.5.1155] [Medline: 16685005]

17.   O'Shea SI, Arcasoy MO, Samsa G, Cummings SE, Thames EH, Surwit RS, et al. Direct-to-patient expert system and home
       INR monitoring improves control of oral anticoagulation. J Thromb Thrombolysis 2008 Aug;26(1):14-21. [doi:
       10.1007/s11239-007-0068-y] [Medline: 17616845]

18.   Veeger NJ, Piersma-Wichers M, Tijssen JG, Hillege HL, van der Meer J. Individual time within target range in patients
       treated with vitamin K antagonists: main determinant of quality of anticoagulation and predictor of clinical outcome. A
       retrospective study of 2300 consecutive patients with venous thromboembolism. Br J Haematol 2005 Feb;128(4):513-519.
       [doi: 10.1111/j.1365-2141.2004.05348.x] [Medline: 15686461]

19.   Connolly SJ, Ezekowitz MD, Yusuf S, Eikelboom J, Oldgren J, Parekh A, RE-LY Steering Committee and Investigators.
       Dabigatran versus warfarin in patients with atrial fibrillation. N Engl J Med 2009 Sep 17;361(12):1139-1151. [doi:
       10.1056/NEJMoa0905561] [Medline: 19717844]

20.   Granger CB, Alexander JH, McMurray JJ, Lopes RD, Hylek EM, Hanna M, ARISTOTLE Committees and Investigators.
       Apixaban versus warfarin in patients with atrial fibrillation. N Engl J Med 2011 Sep 15;365(11):981-992. [doi:
       10.1056/NEJMoa1107039] [Medline: 21870978]

21.   Patel MR, Mahaffey KW, Garg J, Pan G, Singer DE, Hacke W, ROCKET AF Investigators. Rivaroxaban versus warfarin
       in nonvalvular atrial fibrillation. N Engl J Med 2011 Sep 8;365(10):883-891. [doi: 10.1056/NEJMoa1009638] [Medline:
       21830957]

22.   Wan Y, Heneghan C, Perera R, Roberts N, Hollowell J, Glasziou P, et al. Anticoagulation control and prediction of adverse
       events in patients with atrial fibrillation: a systematic review. Circ Cardiovasc Qual Outcomes 2008 Nov;1(2):84-91 [FREE
       Full text] [doi: 10.1161/CIRCOUTCOMES.108.796185] [Medline: 20031794]

23.   Körtke H, Minami K, Breymann T, Seifert D, Baraktaris A, Wagner O, et al. [INR self-management after mechanical heart
       valve replacement: ESCAT (Early Self-Controlled Anticoagulation Trial)]. Z Kardiol 2001;90 Suppl 6:118-124. [Medline:
       11826814]

24.   Christensen TD, Maegaard M, Sørensen HT, Hjortdal VE, Hasenkam JM. Self-management versus conventional management
       of oral anticoagulant therapy: A randomized, controlled trial. Eur J Intern Med 2006 Jul;17(4):260-266. [doi:
       10.1016/j.ejim.2005.11.021] [Medline: 16762775]

25.   Matchar DB, Jacobson A, Dolor R, Edson R, Uyeda L, Phibbs CS, THINRS Executive Committee and Site Investigators.
       Effect of home testing of international normalized ratio on clinical events. N Engl J Med 2010 Oct 21;363(17):1608-1620.
       [doi: 10.1056/NEJMoa1002617] [Medline: 20961244]

26.   Menéndez-Jándula B, Souto JC, Oliver A, Montserrat I, Quintana M, Gich I, et al. Comparing self-management of oral
       anticoagulant therapy with clinic management: a randomized trial. Ann Intern Med 2005 Jan 4;142(1):1-10. [Medline:
       15630104]

27.   Siebenhofer A, Rakovac I, Kleespies C, Piso B, Didjurgeit U, SPOG 60+ Study Group. Self-management of oral
       anticoagulation reduces major outcomes in the elderly. A randomized controlled trial. Thromb Haemost 2008
       Dec;100(6):1089-1098. [Medline: 19132235]

28.   Ansell JE. Empowering patients to monitor and manage oral anticoagulation therapy. JAMA 1999 Jan 13;281(2):182-183.
       [Medline: 9917124]

29.   Latham CE. Is there data to support the concept that educated, empowered patients have better outcomes? J Am Soc Nephrol
       1998 Dec;9(12 Suppl):S141-S144. [Medline: 11443762]

30.   Kimmel SE, Chen Z, Price M, Parker CS, Metlay JP, Christie JD, et al. The influence of patient adherence on anticoagulation
       control with warfarin: results from the International Normalized Ratio Adherence and Genetics (IN-RANGE) Study. Arch
       Intern Med 2007 Feb 12;167(3):229-235. [doi: 10.1001/archinte.167.3.229] [Medline: 17296877]

31.   Horstkotte D, Piper C, Wiemer M. Optimal Frequency of Patient Monitoring and Intensity of Oral Anticoagulation Therapy
       in Valvular Heart Disease. J Thromb Thrombolysis 1998 Jan;5 Suppl 1(3):19-24. [Medline: 10767128]

32.   Heneghan C, Alonso-Coello P, Garcia-Alamino JM, Perera R, Meats E, Glasziou P. Self-monitoring of oral anticoagulation:
       a systematic review and meta-analysis. Lancet 2006 Feb 4;367(9508):404-411. [doi: 10.1016/S0140-6736(06)68139-7]
       [Medline: 16458764]

33.   Samsa GP, Matchar DB. Relationship between test frequency and outcomes of anticoagulation: a literature review and
       commentary with implications for the design of randomized trials of patient self-management. J Thromb Thrombolysis
       2000 Apr;9(3):283-292. [Medline: 10728029]

34.   Beyth RJ, Quinn L, Landefeld CS. A multicomponent intervention to prevent major bleeding complications in older patients
       receiving warfarin. A randomized, controlled trial. Ann Intern Med 2000 Nov 7;133(9):687-695. [Medline: 11074901]

35.   Schulman S, Parpia S, Stewart C, Rudd-Scott L, Julian JA, Levine M. Warfarin dose assessment every 4 weeks versus
       every 12 weeks in patients with stable international normalized ratios: a randomized trial. Ann Intern Med 2011 Nov
       15;155(10):653-9, W201. [doi: 10.1059/0003-4819-155-10-201111150-00003] [Medline: 22084331]

36.   Regier DA, Sunderji R, Lynd LD, Gin K, Marra CA. Cost-effectiveness of self-managed versus physician-managed oral
       anticoagulation therapy. CMAJ 2006 Jun 20;174(13):1847-1852 [FREE Full text] [doi: 10.1503/cmaj.051104] [Medline:
       16785459]

37. Ryan F, Byrne S, O'Shea S. Randomized controlled trial of supervised patient self-testing of warfarin therapy using an internet-based expert system. J Thromb Haemost 2009 Aug;7(8):1284-1290. [doi: 10.1111/j.1538-7836.2009.03497.x] [Medline: 19496921]

38. Harper P, Pollock D. Improved anticoagulant control in patients using home international normalized ratio testing and decision support provided through the Internet. Intern Med J 2011 Apr;41(4):332-337. [doi: 10.1111/j.1445-5994.2010.02282.x] [Medline: 20546057]

39. Ogilvie IM, Newton N, Welner SA, Cowell W, Lip GY. Underuse of oral anticoagulants in atrial fibrillation: a systematic review. Am J Med 2010 Jul;123(7):638-645.e4. [doi: 10.1016/j.amjmed.2009.11.025] [Medline: 20609686]

40. Murray E, Fitzmaurice D, McCahon D, Fuller C, Sandhur H. Training for patients in a randomised controlled trial of self management of warfarin treatment. BMJ 2004 Feb 21;328(7437):437-438 [FREE Full text] [doi: 10.1136/bmj.328.7437.437] [Medline: 14976099]

Abbreviations

INR: international normalized ratio
POC: point-of-care
PST: patient self-testing
TTR: time in therapeutic range
UC: usual care

Review

Online Prevention Aimed at Lifestyle Behaviors: A Systematic Review of Reviews

Leonie FM Kohl[1], MSc; Rik Crutzen[1], PhD; Nanne K de Vries[1], PhD

Department of Health Promotion, CAPHRI School for Public Health and Primary Care, Maastricht University, Maastricht, Netherlands

Corresponding Author:
Rik Crutzen, PhD
Department of Health Promotion
CAPHRI School for Public Health and Primary Care
Maastricht University
PO Box 616
Maastricht, 6200 MD
Netherlands
Phone: 31 43 388 28 28
Fax: 31 43 367 10 32
Email: rik.crutzen@maastrichtuniversity.nl

Abstract

Background: Interventions aimed at behavior change are increasingly being delivered over the Internet. Although research on intervention effectiveness has been widely conducted, their true public health impact as indicated by reach, effectiveness, and use is unclear.

Objective: The aim of this paper is to (1) review the current literature on online prevention aimed at lifestyle behaviors, and (2) identify research gaps regarding reach, effectiveness, and use.

Methods: A systematic search in PubMed revealed relevant literature published between 2005 and 2012 on Internet-delivered behavior change interventions aimed at dietary behaviors, physical activity, alcohol use, smoking, and condom use. Our search yielded 41 eligible reviews, which were analyzed in terms of reach, effectiveness, and use according to the RE-AIM framework.

Results: According to health priorities, interventions are largely targeted at weight-related behaviors, such as physical activity and dietary behavior. Evaluations are predominantly effect-focused and overall effects are small, variable, and not sustainable. Determinants of effectiveness are unclear; effectiveness cannot yet be unambiguously attributed to isolated elements. Actual reach of interventions is undiversified, mostly reaching participants who are female, highly educated, white, and living in high-income countries. One of the most substantial problems in online prevention is the low use of the interventions, a phenomenon seen across all behavior domains.

Conclusions: More research is needed on effective elements instead of effective interventions, with special attention to long-term effectiveness. The reach and use of interventions need more scientific input to increase the public health impact of Internet-delivered interventions.

Introduction

Unhealthy lifestyles have a major impact on morbidity and mortality [1]. Health promotion is the process of enabling people to increase control over their health and its determinants, thereby improving their health [2]. This also entails interventions targeting lifestyle behaviors. Health promotion is shifting its gaze toward new delivery modes (eg, the Internet) to effectively reach a larger part of the population with interventions targeting lifestyle behaviors. Access to the Internet is growing, especially in high-income areas, such as the United States (78.6%) and Europe (63.2%) [3]. Therefore, Internet-delivered interventions have become a well-established instrument within the health promoter's toolbox to potentially reach a large part of the population. Internet-delivered interventions can be operationalized as "typically behaviorally or cognitive behaviorally-based treatments that have been operationalized

and transformed for delivery via the Internet". Usually, they are highly structured; self or semi self-guided; based on effective face-to-face interventions; personalized to the user; interactive; enhanced by graphics, animations, audio, and possibly video; and tailored to provide follow-up and feedback" [4].

The advantages of using the Internet as a delivery mode for health promotion are numerous. From a users' point of view, the Internet is accessible 24/7 and interventions can be used anonymously and at any pace [5]. Anonymity is favorable for interventions regarding behaviors that might involve shame (eg, condom use, alcohol moderation) [6]. Internet-delivered interventions may reach populations who otherwise would not receive treatment [7]. From a health promoters' point of view, (1) the Internet offers improved opportunities for maintenance and updating of interventions [8], (2) Internet-delivered interventions can mimic expensive face-to-face sessions in providing an individual as well as interactive approach in combination with an increased reach [9,10], and (3) Internet-delivered interventions are adjustable relatively easily to the needs of the user [11]. The advantageous characteristics of Internet-delivered interventions are also related to the downsides. Anonymity and limited face-to-face contact, for instance, can lead to high attrition rates [12]. This might affect the public health impact of these interventions.

Still, Internet-delivered interventions seem to possess potential with respect to health behavior change [13], but the actual public health impact remains unknown. Within the field of public health, intervention research is predominantly focused on intervention effectiveness [14], or even on effective elements [15]. This research focus oversimplifies reality in the quest to identify efficacious interventions. Effectiveness alone may reflect internal validity, but many interventions that prove to be effective in trials are much less effective when disseminated outside the context of a trial [16]. It has become apparent that reach and use of interventions are at least as important as effectiveness, because the most effective intervention will not have a public health impact if its reach or actual use by the target group is low [17].

The RE-AIM framework acknowledges that reach, effectiveness, adoption, implementation, and maintenance (hence, the acronym RE-AIM) are factors that all contribute to the public health impact of an intervention [18]. Therefore, this study focuses on the public health impact of Internet-delivered interventions by taking all these factors into account [19]. In the RE-AIM framework, reach is described as the percentage of individuals affected by a policy or program, but also the characteristics of those individuals [19]. In this review, effectiveness is defined as changes in behavioral outcomes. Originally, the RE-AIM framework described adoption, implementation, and maintenance as factors at the organizational level. Internet-delivered interventions, however, can often be used standalone, which means that adoption, implementation, and maintenance also takes place (at least partly) at the individual level: each user decides whether to visit an intervention website for the first time (adoption), and whether to keep using it as intended (implementation), and for the long term (maintenance) [20,21]. These factors of the RE-AIM model are defined as *use* within the current study.

Within the field of public health, the use of the Internet as (the primary) delivery mode has expanded substantially and it is hard to imagine the public health field without the Internet. However, as mentioned previously, the public health impact as indicated by reach, effectiveness, and use remains unknown. This literature study comprises a systematic review of reviews addressing the following research questions:

1. What is the reach, effectiveness, and use of Internet-delivered interventions aimed at lifestyle behaviors (ie, dietary behaviors, physical activity, alcohol use, smoking, and condom use)?
2. What are the gaps in our current knowledge about the public health impact of Internet-delivered interventions aimed at lifestyle behaviors?

Methods

Search Strategy

We identified relevant publications by conducting a PubMed search. The search query was designed in a way that both a search word regarding one of the behaviors of interest was in the title or abstract as well as a search word indicating the use of the Internet as (primary) delivery mode. Terms (including spelling variations and synonyms) that we searched for were eHealth OR Internet AND physical activity (exercise, sport, exertion, training, energy balance), smoking (tobacco, cigarette), alcohol (drinking, AOD, substance), nutrition (food, eat, weight, obesity, overweight, diet, adiposity), sexuality (safe sex, condom, HIV, AIDS, STI, STD), OR behavior (health, lifestyle, prevention, intervention). The exact search query can be found in Multimedia Appendix 1.

The search was conducted in December 2012 and was limited to systematic reviews and meta-analyses published from 2005 to 2012 in the English language. Prior research was covered by a study conducted in 2005 by De Nooijer et al [22] in which no reviews were available.

Selection Criteria

Systematic reviews and meta-analyses were included if they (1) described at least 2 primarily Internet-delivered interventions aimed at behavior change regarding physical activity, smoking, alcohol use, dietary behaviors, or condom use, (2) reported on reach, effectiveness, or use of the included interventions, (3) were aimed at primary or secondary prevention for (part of) the general population, and (4) were available (full text) in English. Interventions aimed at health care workers or other intermediates were excluded. The article selection as well as data extraction was for all studies independently performed by 2 researchers (LK and RC) employing a conservative approach. If 1 of the 2 researchers was in doubt based on the title, the article was taken to the next round assessing the abstract. If 1 of the 2 researchers was in doubt based on the abstract, the article was taken to the next round assessing the full text. Disagreements were discussed until full consensus was reached.

Search Outcome

The database search resulted in 4868 articles. Initial review was based on the title, after which 276 articles remained. Based on

abstracts, 206 articles were excluded leaving 70 articles for a full-text analysis. Reasons for exclusion of articles based on full text (n=30) were that the publication was not a systematic review or meta-analysis (10/30, 33%), contained less than 2 eligible interventions (9/30, 30%), did not report measures on reach, effectiveness, or use of the interventions reviewed (5/30, 17%), were not primarily Internet-delivered (4/30, 13%), were aimed at treatment of a disease (1/30, 3%), or the full text was not available in the English language (1/30, 3%). Some articles were excluded for more than 1 reason. One article was retrieved in a manual search. Reference lists of the selected articles were checked for possible missed publications, but yielded no additional articles. Forty-one articles were found to be eligible for this review (Figure 1).

An additional search was conducted in May 2013. This search extended the initial search by including the terms mhealth, smartphone, and mobile. In the initial search, these terms were not included because the focus of the study was on interventions that are primarily Internet-delivered. Therefore, reviews focusing on text messaging or apps were not included. Internet-delivered interventions are delivered by means of websites and it might indeed be that these are delivered by means of websites especially designed to be used/accessed on mobile phones. The additional search resulted in 174 hits (using the same limitations as the initial search). After title and abstract selection, there were 7 articles appropriate for a full-text analysis. Five of these concerned interventions that were not primarily Internet-delivered (eg, only text messaging). Two articles remained [23,24], but these were already included in the initial search.

Data Extraction

Data were extracted on the target group, number of eligible articles, and the size of corresponding study samples. Our main interest was what the studies reported on the reach, effectiveness, and use of the interventions reviewed. Reported limitations and recommendations were extracted also. Some of the selected reviews partly contained studies that did not correspond with the objectives of this study (eg, non–Internet-delivered interventions). Only data reflecting the eligible interventions have been reported and these data are included in Multimedia Appendix 2.

Quality Assessment

A quality assessment was performed using the AMSTAR tool as well as the more detailed and sensitive R-AMSTAR tool. These tools are especially designed to assess the quality of systematic reviews and meta-analyses. With the AMSTAR tool, such studies are assessed on 11 quality criteria (yes/no/can't answer/not applicable); total scores can range from 0 to 11 [25]. The revised AMSTAR tool (R-AMSTAR) uses the same 11 criteria, but subdivides them into separate items, making the R-AMSTAR more sensitive. Total R-AMSTAR scores can range from 11 to 44 [26]. Both these tools have been validated [26-28].

Figure 1. Flowchart of study selection process.
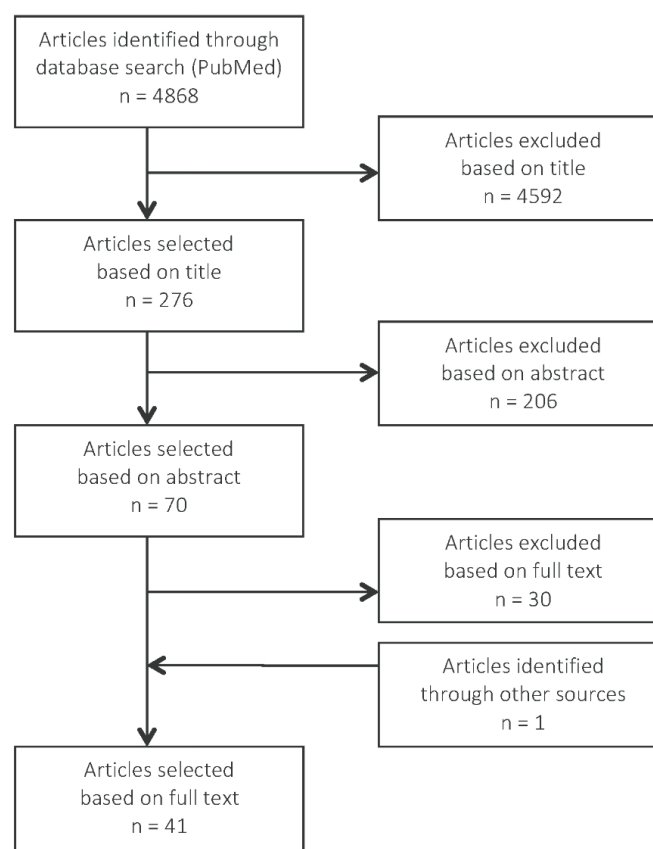
Results

Characteristics of the Selected Studies

We included 41 papers [13,15,23,24,29-65] after article selection (Multimedia Appendix 2). Most studies were systematic reviews (27/41, 66%) and 16 reported meta-analyses (16/41, 39%) [13,15,30,33,35-41,43-45,52,59]. The studies were classified into 8 groups according to the behavior they targeted. The largest group targeted weight management, comprising 11 studies aimed at both dietary behaviors as well as physical activity [24,29,31,37,38,40,42,47-49,64]. All these were aimed at weight loss, and there were 5 reviews that also included interventions on weight maintenance [31,37,40,47,49]. Six studies included 3 or more behaviors [13,15,54,56-58]. The other groups included studies aimed at physical activity (6/41, 15%) [23,36,51,53,60,65], smoking (5/41, 12%) [30,35,39,45,46], alcohol use (5/41, 12%) [33,43,55,61,62], substance use including combined alcohol use and smoking (4/41, 10%) [34,44,52,63], and dietary behaviors (3/41, 7.3%) [32,50,59]. An additional manual search revealed a study on condom use (1/41, 2.4%) [41]. All studies were published between 2006 and 2012. Because we had broad inclusion criteria, study populations ranged from children aged 6 years [23] to people aged 50 years and older [29]. Most reviews were aimed at adults; 30 of 38 reviews (79%) reported age groups including adult populations. The mean of the AMSTAR score on overall quality of the included studies was 3.56 (SD 2.06). The mean of the R-AMSTAR score was 25.5 (SD 5.20). Item scores for all included studies are available in Multimedia Appendix 3.

Reach

Most of the time, a very homogeneous sample of the population is reached. There is a strong female dominance, especially in weight loss/management interventions. Of the 18 reviews that reported gender distribution of the total sample, 16 reviews (89%) reported reaching more women than men [29-32,36,39,40,42,47,49,51,55,59,60,62,65]. Participants were predominantly highly educated [13], young, white [30,31], and living in high-income countries [32,33]. They were approached primarily through traditional offline recruitment efforts; however, some studies use online advertising or suggest to use more modern recruitment strategies [29].

Effectiveness

Overall, Internet-delivered interventions seem to have the potential to achieve behavior change. However, effect sizes were small, rarely moderate [13,15,29,33-44,66]. Internet-delivered interventions compared to a no-treatment control condition had larger effect sizes than when compared with other interventions [13,30,45-47]. In some cases, often including face-to-face elements, the control was equally or more effective than the Internet-delivered intervention [24,31,37,42,46]. It is not exactly clear what effective elements were and for whom these were effective [15,40,46]. For some, indications of effectiveness were found. From this review, tailored feedback [13,24,33,43,47,48], use of theory [15,36,38], interactivity [30,38], goal setting [24], and combinations of

online and in-person contact [32] emerge as noteworthy promising constructs. It remains to be explored what elements work in what situation, and in what combination [13,29,40,47-51]. Furthermore, it is also unclear when interventions become cost-effective [15,33,36,37,43-45,47,52,53]. Long-term effects are measured in a limited number of interventions [34,36,38,39,50-52]; these effects are often unknown and the results that are available show very limited sustained effects (≥6 months, following the RE-AIM framework individual maintenance standards). Studies indicate that effect sizes decrease with intervention length and postintervention or that behavior is not maintained at all because effect sizes were quite small initially [38,51,52]. Embedding an Internet-delivered intervention in existing structures (eg, schools, health care) might increase effectiveness [54,55]. Whether increased reach or use are the underlying driving forces behind this increased effectiveness is uncertain.

Use

One of the largest problems in Internet-delivered interventions is low actual use. There is a wide variety of terminology used to describe use-related constructs (eg, adherence, exposure, and intervention attrition). Experimental research and theory development regarding intervention use is still in its infancy. Given the dose-response relationship between use and effectiveness, this is crucial to improve Internet-delivered interventions aimed at inducing behavior change. Factors suggested to stimulate the use of an intervention were sending reminders [40,54,56], incorporating professional support [54,56,57], and embedding interventions in existing structures [54,55]. Process evaluations should explore people's user experience in order to be able to adjust interventions accordingly [58].

An illustrative description of the average Internet-delivered intervention states that a typical specimen is meant to be used once a week, is modular in setup, is updated once a week, lasts for 10 weeks, includes interaction with the system, a counselor, and peers on the Internet, includes some persuasive technology elements, and results in approximately 50% of the participants adhering to the intervention [56]. Lastly, an issue regarding the use of an intervention is that this is mostly not objectively measured, and if so, heterogeneously, very poorly or not at all described [56,57].

Dietary Behaviors

The reviews on improving dietary behaviors primarily focused on younger populations (children, adolescents, young adults) [32,50]. One study performed a cost-effectiveness analysis and concluded that eHealth devices are unlikely to be cost-effective [59]. The studies showed small effects and the limited data on follow-up measures show that these effects were usually not maintained. The number of available reviews was small (3/41, 7.3%), and individual studies including follow-up measures were even scarcer (1/3, 33%). It was also recommended that the role of social support in Internet-delivered interventions shows promise and should be investigated more thoroughly [50].

Table 1. Quality assessment of included studies.

Reference	Behavior	Study design[a]	AMSTAR score[b]	R-AMSTAR score[c]
Harris et al (2011) [59]	Dietary behaviors	M	9	37
Hamel et al (2012) [50]	Dietary behaviors	SR	4	27
Lua et al (2012) [32]	Dietary behaviors	SR	1	18
Davies et al (2012) [36]	Physical activity	M	6	33
Lau et al (2011) [23]	Physical activity	SR	3	24
Laplante et al (2011) [65]	Physical activity	SR	4	26
Hamel et al (2011) [60]	Physical activity	SR	5	29
Van den Berg et al (2007) [53]	Physical activity	SR	2	24
Vandelanotte et al (2007) [51]	Physical activity	SR	1	23
Bewick et al (2008) [61]	Alcohol use	SR	3	25
Riper et al (2009) [43]	Alcohol use	M	5	31
Vernon (2010) [62]	Alcohol use	SR	1	18
White et al (2010) [55]	Alcohol use	SR	3	22
Riper et al (2011) [33]	Alcohol use	M	5	34
Myung et al (2009) [39]	Smoking	M	5	31
Shahab and McEwen (2009) [30]	Smoking	M	2	25
Civjlak et al (2010) [45]	Smoking	M	6	33
Hutton et al (2011) [46]	Smoking	SR	3	27
Chen et al (2012) [35]	Smoking	M	10	38
Noar et al (2009) [41]	Condom use	M	4	28
Weinstein et al (2006) [31]	Weight management	SR	1	18
Saperstein et al (2007) [48]	Weight management	SR	1	17
Maon et al (2012) [38]	Weight management	M	4	24
Fry et al (2009) [49]	Weight management	SR	2	20
An et al (2009) [64]	Weight management	SR	3	21
Neve et al (2010) [40]	Weight management	M	6	31
Manzoni et al (2011) [47]	Weight management	SR	2	18
Aalbers et al (2011) [29]	Weight management	SR	2	25
Kodama et al (2012) [37]	Weight management	M	5	32
Coons et al (2012) [24]	Weight management	SR	2	23
Norman et al (2007) [42]	Weight management	SR	2	22
Champion et al (2012) [34]	Substance use	SR	4	24
Lehto et al (2011) [63]	Substance use	SR	4	26
Rooke et al (2010) [44]	Substance use	M	2	23
Tait et al (2010) [52]	Substance use	M	3	23
Kelders et al (2012) [56]	Multiple behaviors	SR	3	24
Donkin et al (2011) [58]	Multiple behaviors	SR	1	21
Brouwer et al (2011) [57]	Multiple behaviors	SR	2	22
Crutzen et al (2011) [54]	Multiple behaviors	SR	4	27
Cugelman et al (2011) [13]	Multiple behaviors	M	5	27
Webb et al (2010) [15]	Multiple behaviors	M	3	23

Physical Activity

Most reviews on physical activity conclude that when studies are effective, these show modest effect sizes with decreasing effectiveness during follow-up [36,51,53,60]. A large meta-analysis showed that longer intervention duration, the use of social cognitive theory, the use of educational components, and regularly updating the content of the intervention had significant effects on physical activity levels [36]. There was a reoccurring call for research on increasing intervention use [23,36,51].

Alcohol Use

In alcohol abuse prevention, the risk profile of users is thought to be related to the effectiveness of the intervention [61]. It is suggested that brief interventions may be more effective for high-risk participants than low-risk participants. Several studies focused on the use and content of feedback, in which findings on normative and personalized feedback seemed promising, but still inconclusive [43,62]. It was suggested to add a face-to-face component as an adjunct to Internet-delivered interventions to increase effectiveness, but effectiveness is not yet confirmed by research [33]. Use is a problem [43,61]; 1 review showed that women, people married or living with a partner, and those without children were more likely to complete the program [62]. Collecting user feedback to tailor interventions could increase this outcome [62].

Smoking

Almost all reviews aimed at smoking cessation recommended research that focuses on the use of the intervention, which is suggested to be improved by assessing barriers to participation [46]. Effectiveness may be improved by involvement of users in the design of interventions [30,35]. Adolescents form a target group in need of extra attention [35,39,46]. Biochemical measures to measure smoking behavior may be more accurate, but effectiveness appears to remain the same regardless of the method [30,39,46]. Increasing interactivity and message tailoring seemed effective strategies [30,35,46]. In smoking cessation, motivated participants, which were often more females than males, tended to show larger effects on behavior change [30].

Condom Use

There was only 1 systematic review available on condom use, which discussed 3 Internet-delivered interventions [41]. Individualized tailoring and the use of the stages of change model were related to effectiveness, but these were overall conclusions, not specified for Internet-delivered interventions.

Weight Management

In the weight management reviews, the majority of studies reported a strong overrepresentation of women (7/11, 64%), which makes a generalization of the effects to men questionable [31,47,49]. Evidence points toward the effectiveness of including a face-to-face component in Internet-delivered weight loss interventions. When personal contact is part of the intervention, results generally showed that effect sizes and use were higher [31,40,47,48]. When personal contact was not part of the intervention, reverse effects were seen [37]. There was, however, no conclusive evidence. Interactive elements (eg, online peer support or forums) were shown to increase effectiveness; however, use was very limited [29]. Some studies did not find a difference between Internet-delivered and traditional (face-to-face) interventions [24,31,38,40,47]. It was suggested that a difference was hard to find because of the high attrition rates [24] or that this finding might reflect an equal effectiveness, which could mean opportunities to increase interventions' cost-effectiveness [37]. Five reviews (5/11, 45%) included weight maintenance interventions in which the weight maintenance studies were always outnumbered by the weight loss studies [31,37,40,47,49]. The limited findings indicate face-to-face interventions are more effective than Internet-delivered interventions.

Substance Use

Most studies in the reviews on substance use were aimed at adolescents or young adults, especially with regard to cigarette smoking behavior [34,52,63]. It is not clear what elements of the interventions were effective, but suggestions were made toward parental involvement, number of sessions, so-called booster sessions, normative education, resistance skills training, and reducing positive expectancies [34]. Also within these behavior domains there was a demand for research that compares Internet-delivered with in-person interventions or combinations of the 2 [44,52]. One meta-analysis concluded that Internet-delivered interventions were significantly less effective than offline computer-delivered interventions [44].

Multiple Behaviors

As opposed to the other review sections, effectiveness was not always the main outcome when multiple behaviors were considered. Reviews also focused on intervention use, which has been shown to be related to effectiveness. For instance, more adherent participants of weight loss programs lost more weight [58]. Because effectiveness was the most commonly reported outcome, some reviews experienced difficulties collecting their data due to poor reporting on other constructs than effectiveness [54,56,57]. Several reviews showed that nonadherence was a major problem of Internet-delivered interventions in particular, and a large barrier to effective interventions [58]. Frequency of interaction (eg, email and/or phone contact with visitors, but also reminders), dialog support, updates [56], targeted/tailored communication, monitoring of behavior change, professional and peer support, interactive and easy accessible content, conditional progress, incentives, and embedment in a social context all seemed to increase use [54].

With regard to effectiveness, the reviews found that primary task support [56], extensive use of theory, inclusion of more behavior change techniques, elaborate interaction with participants, and training in stress management and general communication skills showed significant results on behavior

change [15]. Intervention duration was negatively correlated with impact [13]. There was low use of interactive elements with peers and with professionals [54].

Discussion

Principal Findings

The sizable growth of Internet-delivered interventions can be recognized when comparing the 41 reviews found in this study to the 9 individual studies found in a study with similar objectives conducted in 2005 [22]. Almost half of the reviews on Internet-delivered interventions (20/41, 49%) were aimed at overweight-related behaviors (physical activity and dietary behaviors), reflecting a research focus congruent with health priorities [67]. For smoking, the leading cause of preventable deaths in the United States [68], 9 reviews were found. For condom use, data on Internet-delivered interventions were very limited; only 1 review was available [41] focusing on computer-delivered interventions in general (ie, not necessarily Internet-delivered).

This review confirms the previously mentioned statement that within the public health field, there is a strong focus on effectiveness. It is reported most often and thoroughly, but effect sizes appear to be small, variable, and behavior change appears to be unsustainable at follow-up measures. Effect sizes may be small, but it should be noted that the Internet is a delivery mode with an unrivaled potential reach and this may still enhance public health impact [69]. Estimates of actual reach for Internet interventions are lacking [17]. The intended reach of Internet-delivered interventions is varied, aiming at a diverse population with respect to gender, socioeconomic status, and ethnic background, but the actual reach is undiversified; mostly the female, Caucasian, highly educated part of the population is reached, corresponding to previous findings [70]. It is also desirable to reach high-risk participants; however, these are not necessarily attracted at the same rate as low-risk participants [71]. Another limitation of the available studies is that most research is predominantly aimed at high-income populations [32,33], which makes generalization to low- and middle-income countries questionable. Although the potential reach of Internet-delivered interventions is virtually infinite, this review shows that, at least in terms of reach, there is still much to be gained.

Effect sizes are small, variable, and not sustainable. But what determines effectiveness? It is often not clear what the relevant active ingredients are for effectiveness [72,73], and there is a clear call for elucidation on this subject from this review. The current data show indicative evidence for the effectiveness of tailored feedback, use of theory, interactivity, goal setting, and combinations of online and in-person contact. For several behavior change techniques, there seems to be some evidence, not all equally conclusive. Some well-substantiated evidence can be found for the effectiveness of message tailoring [74,75], but variations between interventions exist in the operationalization, and effectiveness is not guaranteed in all cases [76]. A large meta-analysis found that extensive use of theory and the use of multiple behavior change techniques predict effectiveness in Internet-delivered interventions [15].

Including face-to-face elements is recommended or found to increase the use or effectiveness in weight loss interventions [31,37,40], alcohol use interventions [33], and smoking cessation interventions [35]. This could, however, have negative implications for reach. Including face-to-face support undermines the potentially high reach of Internet-delivered interventions, because of a high therapist burden. It is recommended to find a balance between face-to-face elements and self-guided Internet-delivered material [30,31,33,35,37], also in the light of cost-effectiveness [37]. Adjunctive designs have been applied to Internet-delivered interventions; however, it is not yet discerned what elements cause the effect of an intervention and whether these adjunctive elements contribute and to what extent [77]. Some reviews indicated that interventions show improved effectiveness when they are embedded in an existing structure, such as schools or health care. This may not only improve effectiveness, but may improve reach and use as well. It is also suggested to use Internet-delivered interventions as a part of a stepped-care model [78]. In this way, Internet-delivered interventions could serve as a first step in which individual needs are assessed with respect to the necessary amount of support, time, and expertise [79].

Internet-delivered interventions are likely to be less costly than a face-to-face intervention and this is an oft-cited reason for delivering an intervention online [11]. This reasoning probably holds for fully automated systems, but a fair point is made when considering that Internet-delivered interventions can still contain a substantial amount of human involvement [79], which makes assumptions on cost-effectiveness less certain. Research on cost-effectiveness was recommended repeatedly in the reviews considered in the current study. A study answering this call concluded that cost-effectiveness is hard to determine, especially due to a lack of data [80].

Data on the interventions' use were poorly reported in most studies; in some cases, these data were completely lacking. These are missed opportunities, because Internet-delivered interventions in particular have the technological advantage to be able to provide more insight into intervention use [58]. From the currently reported data, it is shown that there is particularly low use. The phenomenon defined as the discontinuation of eHealth application use, called *the law of attrition*, is considered to be "one of the fundamental characteristics and methodological challenges in the evaluation of eHealth applications" [12]. In our own findings, the higher attrition rates in Internet-delivered compared to traditional interventions is most clearly illustrated by the review on condom use, in which the 2 trials with the poorest use were delivered online [41]. Participants' nonusage of an intervention can be explained from a reversed diffusion of innovations model [12,81]. The diffusion of innovations model is explained by a symmetric curve depicting the proportion of a population adopting an innovation and their motivations; the reversed model would reflect on the discontinuation of using an innovation (in this case, an intervention). Following this line of reasoning, factors influencing adoption may be used in Internet-delivered interventions to prevent low usage. There are a wide range of factors thought to stimulate intervention use. Christensen and Mackinnon [82] point out the importance of user characteristics

XSL·FO
RenderX

and preferences. A review of adherence in 3 Internet-delivered trials showed a positive association for higher self-efficacy, having less smoking friends, older age, being female, and a higher education with an increased use of the intervention [83], but also interaction with a counselor, more frequent intended usage, more frequent updates, and more extensive employment of dialog support were found to be predictors of improved use [56]. Engaging users is also thought to be improved by the transition from a more static, structured, developer-defined intervention content to less structured, more user-defined interactions [84,85]. Research on the use of Internet-delivered interventions is a relatively young and sometimes overshadowed endeavor; therefore, most findings are not yet confirmed by a substantive body of evidence. It is important that research is contributing to this debate [82], because decreased intervention use has been shown to negatively affect health behavior change [58] and the findings of the current study further underline the relevance of this problem.

Recommendations for Future Research

This review shows that a substantial amount of research has been done, but we found some reoccurring research needs, which will be discussed in terms of reach, effectiveness, and use.

The interventions' reach is found to be undiversified and it is hard to reach high-risk groups. Participant recruitment can be done offline as well as online. Although reach seems to be much larger online, online recruitment can be a disappointing venture, even when targeting a young audience [86]. Recruitment strategies should be aligned with users' search strategies, which have to be studied first. Another strategy to increase reach is to create interest for the intervention. A lack of interest for Internet-delivered lifestyle interventions is identified by Lieberman and Massey [87], who developed a motivational Internet application that was used to increase treatment interest in individuals with drinking problems not receiving treatment. Including a meta-intervention, which can be described as a small intervention prior to the actual intervention, there was increased participation of high-risk participants in an HIV-prevention intervention [88]. It is also found that motivation for behavior change and curiosity regarding the content of the intervention appear to be important for a first visit to an Internet-delivered intervention [89]. An extension of this line of research is warranted.

The largest gap in research on effectiveness seems to be that it is not known what intervention elements are effective and under what conditions. Future research should entail experimental studies focusing on testing isolated ingredients of Internet-delivered interventions. Effect evaluations on complete interventions are widely available and interventions proven to be effective could provide a starting point to disentangle effective ingredients. These may also be found outside the scope of Internet-based strategies, in adjunctive designs. Here, online and offline strategies are combined and it would be very interesting to investigate optimal combinations, possibly also in the context of cost-effectiveness. Interventions should include user profiles with information on a wide range of user characteristics. Such data are not only suitable for developing tailored advice, but could also shed light on effectiveness for

subgroups of the population (eg, high-risk groups) to answer the question: what is effective for whom?

Discontinued or suboptimal use of the intervention is a widely recognized problem, also evident from the current review. There is a strong need for strategies that can be employed to increase use. User involvement in the creation of the intervention may be one of these strategies to keep users engaged. Some work has been done on this topic [89], but these findings need replication and further study. Moreover, there is a need for theory-driven experimental studies focusing on use of Internet-delivered interventions [90].

For all areas of research, reviews recommend better data reporting because inadequate data reporting posits unnecessary limitations to research. This is especially the case when conducting a meta-analysis or review, or in the case of data on the reach or use of interventions. Full disclosure could be a solution, because it stimulates scrutiny and replication of research findings, and improves meta-analytical research [91,92]. Moreover, the current technological developments should be employed more exhaustively. Therefore, cooperation with other disciplines is warranted to be able to develop modern, well-designed interventions. This intensified use of technology could not only be applied to data sharing, but also to increase effectiveness of interventions and to provide more insight into reach and use. Regarding the latter, there are some unused opportunities in terms of data collection [58]. The Internet offers possibilities to collect a wide range of valuable data regarding intervention use (eg, log-ins, page views), which is readily available and should be disclosed.

Limitations

Because of the substantial amount of data, we decided to perform a review of reviews. There are some weaknesses inherent to secondary analyses. It is inevitable that data and details are lost in order to obtain a more robust overview. The quality of our data is directly dependent on the reporting and execution quality of the reviewed studies. We performed a quality assessment on the included studies. The scores approached the lower end of the scales, but were comparable to those found in previous studies [93]. Although the search was executed with the greatest consideration, it is possible that some publications or data may have been missed. This might be because only the primary search engine for this type of research was used to collect our data (ie, PubMed), although we also checked reference lists for completeness of our review, or because publications in English were solely considered in our review. Although we tried to be as elaborate as possible in the methodology, it should be acknowledged that this study may be affected by subjectivity bias because of the nature of the study design (ie, a systematic review of reviews) and quantitative data being available limitedly. Another possible bias lies in the fact that we did not correct for studies appearing in several of the included reviews. The result may be that some of the same evidence is covered more than once. Because we did not perform quantitative analyses, the resulting bias is estimated to be limited.

Despite these possible biases, we believe that the systematically collected data contributes to our understanding and to a general

overview of what research has been done and what research still needs to be done.

Conclusion

This review provides an overview concerning research on online prevention aimed at lifestyle behaviors. The findings of this research show that reviews are effect-oriented, but interventions show small, unsustainable effects on behavior change. Research on reach and use of interventions is less advanced and needs more scientific input.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Search query.

[PDF File (Adobe PDF File), 89KB - jmir_v15i7e146_app1.pdf]

Multimedia Appendix 2

Description of included studies.

[PDF File (Adobe PDF File), 342KB - jmir_v15i7e146_app2.pdf]

Multimedia Appendix 3

Quality assessment of included studies.

[XLSX File (Microsoft Excel File), 55KB - jmir_v15i7e146_app3.xlsx]

References

1. Centers for Disease Control and Prevention. 2012. Chronic diseases and health promotion URL: http://www.cdc.gov/chronicdisease/overview/index.htm [accessed 2013-04-10] [WebCite Cache ID 6FlruroEF]
2. World Health Organization. The Bangkok Charter for health promotion in a globalized world. Health Promot J Austr 2005 Dec;16(3):168-171. [Medline: 16375027]
3. Internet World Stats. 2012. World Internet usage and population statistics URL: http://www.internetworldstats.com/stats.htm [accessed 2013-04-10] [WebCite Cache ID 6Fls714uR]
4. Ritterband LM, Thorndike F. Internet interventions or patient education web sites? J Med Internet Res 2006;8:e18. [doi: 10.2196/jmir.8.3.e18]
5. Napolitano MA, Marcus BH. Targeting and tailoring physical activity information using print and information technologies. Exerc Sport Sci Rev 2002 Jul;30(3):122-128. [Medline: 12150571]
6. Moyer A, Finney JW. Brief interventions for alcohol problems: factors that facilitate implementation. Alcohol Res Health 2004;28(1):44-50. [Medline: 19006991]
7. Spek V, Cuijpers P, Nyklícek I, Riper H, Keyzer J, Pop V. Internet-based cognitive behaviour therapy for symptoms of depression and anxiety: a meta-analysis. Psychol Med 2007 Mar;37(3):319-328. [doi: 10.1017/S0033291706008944] [Medline: 17112400]
8. Atkinson NL, Gold RS. The promise and challenge of eHealth interventions. Am J Health Behav 2002;26(6):494-503. [Medline: 12437024]
9. Neuhauser L, Kreps GL. Rethinking communication in the E-health era. J Health Psychol 2003 Jan;8(1):7-23. [doi: 10.1177/1359105303008001426] [Medline: 22113897]
10. Cummins CO, Evers KE, Johnson JL, Paiva A, Prochaska JO, Prochaska JM. Assessing stage of change and informed decision making for Internet participation in health promotion and disease management. Manag Care Interface 2004 Aug;17(8):27-32. [Medline: 15471108]
11. Griffiths F, Lindenmeyer A, Powell J, Lowe P, Thorogood M. Why are health care interventions delivered over the internet? A systematic review of the published literature. J Med Internet Res 2006;8(2):e10 [FREE Full text] [doi: 10.2196/jmir.8.2.e10] [Medline: 16867965]
12. Eysenbach G. The law of attrition. J Med Internet Res 2005;7(1):e11 [FREE Full text] [doi: 10.2196/jmir.7.1.e11] [Medline: 15829473]

XSL•FO
RenderX

13. 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]

14. Sorensen G, Emmons K, Hunt MK, Johnston D. Implications of the results of community intervention trials. Annu Rev Public Health 1998;19:379-416. [doi: 10.1146/annurev.publhealth.19.1.379] [Medline: 9611625]

15. 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]

16. Glasgow RE, Eakin EG, Toobert DJ. How generalizable are the results of diabetes self-management research? The impact of participation and attrition. Diabetes Educ 1996;22(6):573-574. [Medline: 8970287]

17. 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]

18. Glasgow RE, Vogt TM, Boles SM. Evaluating the public health impact of health promotion interventions: the RE-AIM framework. Am J Public Health 1999 Sep;89(9):1322-1327. [Medline: 10474547]

19. National Cancer Institute. 2012. Measuring the use of the RE-AIM model dimension items URL: http://www.re-aim.org/forms/re-aim_framework_constructs_checklist.pdf [accessed 2013-06-28] [WebCite Cache ID 6Hi7F7eKi]

20. Crutzen R, de Nooijer J, Brouwer W, Oenema A, Brug J, de Vries NK. Internet-delivered interventions aimed at adolescents: a Delphi study on dissemination and exposure. Health Educ Res 2008 Jun;23(3):427-439 [FREE Full text] [doi: 10.1093/her/cym094] [Medline: 18209115]

21. Crutzen R, de Nooijer J, Brouwer W, Oenema A, Brug J, de Vries NK. A conceptual framework for understanding and improving adolescents' exposure to Internet-delivered interventions. Health Promot Int 2009 Sep;24(3):277-284 [FREE Full text] [doi: 10.1093/heapro/dap018] [Medline: 19515716]

22. De Nooijer J, Oenema A, Kloek G, Brug H, De Vries H, De Vries N. Bevordering van gezond gedrag via internet, nu en in de toekomst [Promoting healthy behavior over the Internet, now and in the future]. Maastricht: Maastricht University; 2005.

23. Lau PW, Lau EY, Wong del P, Ransdell L. A systematic review of information and communication technology-based interventions for promoting physical activity behavior change in children and adolescents. J Med Internet Res 2011;13(3):e48 [FREE Full text] [doi: 10.2196/jmir.1533] [Medline: 21749967]

24. Coons MJ, Demott A, Buscemi J, Duncan JM, Pellegrini CA, Steglitz J, et al. Technology interventions to curb obesity: A systematic review of the current literature. Curr Cardiovasc Risk Rep 2012 Apr;6(2):120-134. [doi: 10.1007/s12170-012-0222-8] [Medline: 23082235]

25. Shea BJ, Grimshaw JM, Wells GA, Boers M, Andersson N, Hamel C, et al. Development of AMSTAR: a measurement tool to assess the methodological quality of systematic reviews. BMC Med Res Methodol 2007;7:10 [FREE Full text] [doi: 10.1186/1471-2288-7-10] [Medline: 17302989]

26. Kung J, Chiappelli F, Cajulis OO, Avezova R, Kossan G, Chew L, et al. From systematic reviews to clinical recommendations for evidence-based health care: Validation of revised assessment of multiple systematic reviews (R-AMSTAR) for grading of clinical relevance. Open Dent J 2010;4:84-91 [FREE Full text] [doi: 10.2174/1874210601004020084] [Medline: 21088686]

27. Shea BJ, Bouter LM, Peterson J, Boers M, Andersson N, Ortiz Z, et al. External validation of a measurement tool to assess systematic reviews (AMSTAR). PLoS One 2007;2(12):e1350 [FREE Full text] [doi: 10.1371/journal.pone.0001350] [Medline: 18159233]

28. Shea BJ, Hamel C, Wells GA, Bouter LM, Kristjansson E, Grimshaw J, et al. AMSTAR is a reliable and valid measurement tool to assess the methodological quality of systematic reviews. J Clin Epidemiol 2009 Oct;62(10):1013-1020. [doi: 10.1016/j.jclinepi.2008.10.009] [Medline: 19230606]

29. Aalbers T, Baars MA, Rikkert MG. Characteristics of effective Internet-mediated interventions to change lifestyle in people aged 50 and older: a systematic review. Ageing Res Rev 2011 Sep;10(4):487-497. [doi: 10.1016/j.arr.2011.05.001] [Medline: 21628005]

30. Shahab L, McEwen A. Online support for smoking cessation: a systematic review of the literature. Addiction 2009 Nov;104(11):1792-1804. [doi: 10.1111/j.1360-0443.2009.02710.x] [Medline: 19832783]

31. Weinstein PK. A review of weight loss programs delivered via the Internet. J Cardiovasc Nurs 2006;21(4):251-8; quiz 259. [Medline: 16823276]

32. Lua PL, Wan Putri Elena WD. The impact of nutrition education interventions on the dietary habits of college students in developed nations: a brief review. Malays J Med Sci 2012 Jan;19(1):4-14 [FREE Full text] [Medline: 22977369]

33. Riper H, Spek V, Boon B, Conijn B, Kramer J, Martin-Abello K, et al. Effectiveness of E-self-help interventions for curbing adult problem drinking: a meta-analysis. J Med Internet Res 2011;13(2):e42 [FREE Full text] [doi: 10.2196/jmir.1691] [Medline: 21719411]

34. Champion KE, Newton NC, Barrett EL, Teesson M. A systematic review of school-based alcohol and other drug prevention programs facilitated by computers or the Internet. Drug Alcohol Rev 2013 Mar;32(2):115-123. [doi: 10.1111/j.1465-3362.2012.00517.x] [Medline: 23039085]

XSL•FO
RenderX

35. Chen YF, Madan J, Welton N, Yahaya I, Aveyard P, Bauld L, et al. Effectiveness and cost-effectiveness of computer and other electronic aids for smoking cessation: a systematic review and network meta-analysis. Health Technol Assess 2012;16(38):1-205 [FREE Full text] [doi: 10.3310/hta16380] [Medline: 23046909]

36. Davies CA, Spence JC, Vandelanotte C, Caperchione CM, Mummery WK. Meta-analysis of internet-delivered interventions to increase physical activity levels. Int J Behav Nutr Phys Act 2012;9:52 [FREE Full text] [doi: 10.1186/1479-5868-9-52] [Medline: 22546283]

37. Kodama S, Saito K, Tanaka S, Horikawa C, Fujiwara K, Hirasawa R, et al. Effect of Web-based lifestyle modification on weight control: a meta-analysis. Int J Obes (Lond) 2012 May;36(5):675-685. [doi: 10.1038/ijo.2011.121] [Medline: 21694698]

38. Maon S, Edirippulige S, Ware R, Batch J. The use of web-based interventions to prevent excessive weight gain. J Telemed Telecare 2012 Jan;18(1):37-41. [doi: 10.1258/jtt.2011.110306] [Medline: 22101608]

39. Myung SK, McDonnell DD, Kazinets G, Seo HG, Moskowitz JM. Effects of Web- and computer-based smoking cessation programs: meta-analysis of randomized controlled trials. Arch Intern Med 2009 May 25;169(10):929-937. [doi: 10.1001/archinternmed.2009.109] [Medline: 19468084]

40. Neve M, Morgan PJ, Jones PR, Collins CE. Effectiveness of web-based interventions in achieving weight loss and weight loss maintenance in overweight and obese adults: a systematic review with meta-analysis. Obes Rev 2010 Apr;11(4):306-321. [doi: 10.1111/j.1467-789X.2009.00646.x] [Medline: 19754633]

41. Noar SM, Black HG, Pierce LB. Efficacy of computer technology-based HIV prevention interventions: a meta-analysis. AIDS 2009 Jan 2;23(1):107-115. [doi: 10.1097/QAD.0b013e32831c5500] [Medline: 19050392]

42. 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]

43. Riper H, van Straten A, Keuken M, Smit F, Schippers G, Cuijpers P. Curbing problem drinking with personalized-feedback interventions: a meta-analysis. Am J Prev Med 2009 Mar;36(3):247-255. [doi: 10.1016/j.amepre.2008.10.016] [Medline: 19215850]

44. Rooke S, Thorsteinsson E, Karpin A, Copeland J, Allsop D. Computer-delivered interventions for alcohol and tobacco use: a meta-analysis. Addiction 2010 Aug;105(8):1381-1390. [doi: 10.1111/j.1360-0443.2010.02975.x] [Medline: 20528806]

45. Civljak M, Sheikh A, Stead LF, Car J. Internet-based interventions for smoking cessation. Cochrane Database Syst Rev 2010(9):CD007078. [doi: 10.1002/14651858.CD007078.pub3] [Medline: 20824856]

46. Hutton HE, Wilson LM, Apelberg BJ, Tang EA, Odelola O, Bass EB, et al. A systematic review of randomized controlled trials: Web-based interventions for smoking cessation among adolescents, college students, and adults. Nicotine Tob Res 2011 Apr;13(4):227-238. [doi: 10.1093/ntr/ntq252] [Medline: 21350042]

47. Manzoni GM, Pagnini F, Corti S, Molinari E, Castelnuovo G. Internet-based behavioral interventions for obesity: an updated systematic review. Clin Pract Epidemiol Ment Health 2011;7:19-28 [FREE Full text] [doi: 10.2174/1745017901107010019] [Medline: 21552423]

48. Saperstein SL, Atkinson NL, Gold RS. The impact of Internet use for weight loss. Obes Rev 2007 Sep;8(5):459-465. [doi: 10.1111/j.1467-789X.2007.00374.x] [Medline: 17716303]

49. Fry JP, Neff RA. Periodic prompts and reminders in health promotion and health behavior interventions: systematic review. J Med Internet Res 2009;11(2):e16 [FREE Full text] [doi: 10.2196/jmir.1138] [Medline: 19632970]

50. Hamel LM, Robbins LB. Computer- and web-based interventions to promote healthy eating among children and adolescents: a systematic review. J Adv Nurs 2013 Jan;69(1):16-30. [doi: 10.1111/j.1365-2648.2012.06086.x] [Medline: 22757605]

51. Vandelanotte C, Spathonis KM, Eakin EG, Owen N. Website-delivered physical activity interventions a review of the literature. Am J Prev Med 2007 Jul;33(1):54-64. [doi: 10.1016/j.amepre.2007.02.041] [Medline: 17572313]

52. Tait RJ, Christensen H. Internet-based interventions for young people with problematic substance use: a systematic review. Med J Aust 2010 Jun 7;192(11 Suppl):S15-S21. [Medline: 20528701]

53. van den Berg MH, Schoones JW, Vliet Vlieland TP. 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]

54. Crutzen R, de Nooijer J, Brouwer W, Oenema A, Brug J, de Vries NK. Strategies to facilitate exposure to internet-delivered health behavior change interventions aimed at adolescents or young adults: a systematic review. Health Educ Behav 2011 Feb;38(1):49-62. [doi: 10.1177/1090198110372878] [Medline: 21189422]

55. White A, Kavanagh D, Stallman H, Klein B, Kay-Lambkin F, Proudfoot J, et al. Online alcohol interventions: a systematic review. J Med Internet Res 2010;12(5):e62 [FREE Full text] [doi: 10.2196/jmir.1479] [Medline: 21169175]

56. Kelders SM, Kok RN, Ossebaard HC, Van Gemert-Pijnen JE. Persuasive system design does matter: a systematic review of adherence to web-based interventions. J Med Internet Res 2012;14(6):e152 [FREE Full text] [doi: 10.2196/jmir.2104] [Medline: 23151820]

57. Brouwer W, Kroeze W, Crutzen R, de Nooijer J, de Vries NK, Brug J, et al. Which intervention characteristics are related to more exposure to internet-delivered healthy lifestyle promotion interventions? A systematic review. J Med Internet Res 2011;13(1):e2 [FREE Full text] [doi: 10.2196/jmir.1639] [Medline: 21212045]

58.   Donkin L, Christensen H, Naismith SL, Neal B, Hickie IB, Glozier N. A systematic review of the impact of adherence on the effectiveness of e-therapies. J Med Internet Res 2011;13(3):e52 [FREE Full text] [doi: 10.2196/jmir.1772] [Medline: 21821503]

59.   Harris J, Felix L, Miners A, Murray E, Michie S, Ferguson E, et al. Adaptive e-learning to improve dietary behaviour: a systematic review and cost-effectiveness analysis. Health Technol Assess 2011 Oct;15(37):1-160 [FREE Full text] [doi: 10.3310/hta15370] [Medline: 22030014]

60.   Hamel LM, Robbins LB, Wilbur J. Computer- and web-based interventions to increase preadolescent and adolescent physical activity: a systematic review. J Adv Nurs 2011 Feb;67(2):251-268. [doi: 10.1111/j.1365-2648.2010.05493.x] [Medline: 21198800]

61.   Bewick BM, Trusler K, Barkham M, Hill AJ, Cahill J, Mulhern B. The effectiveness of web-based interventions designed to decrease alcohol consumption--a systematic review. Prev Med 2008 Jul;47(1):17-26. [doi: 10.1016/j.ypmed.2008.01.005] [Medline: 18302970]

62.   Vernon ML. A review of computer-based alcohol problem services designed for the general public. J Subst Abuse Treat 2010 Apr;38(3):203-211 [FREE Full text] [doi: 10.1016/j.jsat.2009.11.001] [Medline: 20015607]

63.   Lehto T, Oinas-Kukkonen H. Persuasive features in web-based alcohol and smoking interventions: a systematic review of the literature. J Med Internet Res 2011;13(3):e46 [FREE Full text] [doi: 10.2196/jmir.1559] [Medline: 21795238]

64.   An JY, Hayman LL, Park YS, Dusaj TK, Ayres CG. Web-based weight management programs for children and adolescents: a systematic review of randomized controlled trial studies. ANS Adv Nurs Sci 2009;32(3):222-240. [doi: 10.1097/ANS.0b013e3181b0d6ef] [Medline: 19707091]

65.   Laplante C, Peng W. A systematic review of e-health interventions for physical activity: an analysis of study design, intervention characteristics, and outcomes. Telemed J E Health 2011 Sep;17(7):509-523. [doi: 10.1089/tmj.2011.0013] [Medline: 21718092]

66.   Cavallo DN, Tate DF, Ries AV, Brown JD, DeVellis RF, Ammerman AS. A social media-based physical activity intervention: a randomized controlled trial. Am J Prev Med 2012 Nov;43(5):527-532. [doi: 10.1016/j.amepre.2012.07.019] [Medline: 23079176]

67.   Finkelstein EA, Khavjou OA, Thompson H, Trogdon JG, Pan L, Sherry B, et al. Obesity and severe obesity forecasts through 2030. Am J Prev Med 2012 Jun;42(6):563-570. [doi: 10.1016/j.amepre.2011.10.026] [Medline: 22608371]

68.   Danaei G, Ding EL, Mozaffarian D, Taylor B, Rehm J, Murray CJ, et al. The preventable causes of death in the United States: comparative risk assessment of dietary, lifestyle, and metabolic risk factors. PLoS Med 2009 Apr 28;6(4):e1000058 [FREE Full text] [doi: 10.1371/journal.pmed.1000058] [Medline: 19399161]

69.   Glasgow RE, McKay HG, Piette JD, Reynolds KD. The RE-AIM framework for evaluating interventions: what can it tell us about approaches to chronic illness management? Patient Educ Couns 2001 Aug;44(2):119-127. [Medline: 11479052]

70.   Brouwer W, Oenema A, Raat H, Crutzen R, de Nooijer J, de Vries NK, et al. Characteristics of visitors and revisitors to an Internet-delivered computer-tailored lifestyle intervention implemented for use by the general public. Health Educ Res 2010 Aug;25(4):585-595 [FREE Full text] [doi: 10.1093/her/cyp063] [Medline: 19897515]

71.   Glasgow RE, Nelson CC, Kearney KA, Reid R, Ritzwoller DP, Strecher VJ, et al. Reach, engagement, and retention in an Internet-based weight loss program in a multi-site randomized controlled trial. J Med Internet Res 2007;9(2):e11 [FREE Full text] [doi: 10.2196/jmir.9.2.e11] [Medline: 17513282]

72.   Collins LM, Murphy SA, Nair VN, Strecher VJ. A strategy for optimizing and evaluating behavioral interventions. Ann Behav Med 2005 Aug;30(1):65-73. [doi: 10.1207/s15324796abm3001_8] [Medline: 16097907]

73.   Norman GJ. Answering the "what works?" question in health behavior change. Am J Prev Med 2008 May;34(5):449-450 [FREE Full text] [doi: 10.1016/j.amepre.2008.02.005] [Medline: 18407014]

74.   Noar SM, Benac CN, Harris MS. Does tailoring matter? Meta-analytic review of tailored print health behavior change interventions. Psychol Bull 2007 Jul;133(4):673-693. [doi: 10.1037/0033-2909.133.4.673] [Medline: 17592961]

75.   Kreuter MW, Farrell DW, Olevitch LR, Brennan LK. Tailoring Health Messages: Customizing Communication with Computer Technology. Mahwah, NJ: Lawrence Erlbaum; 1999.

76.   Kreuter MW, Oswald DL, Bull FC, Clark EM. Are tailored health education materials always more effective than non-tailored materials? Health Educ Res 2000 Jun;15(3):305-315 [FREE Full text] [Medline: 10977378]

77.   Danaher BG, Seeley JR. Methodological issues in research on web-based behavioral interventions. Ann Behav Med 2009 Aug;38(1):28-39. [doi: 10.1007/s12160-009-9129-0] [Medline: 19806416]

78.   Ritterband LM, Tate DF. The science of internet interventions. Introduction. Ann Behav Med 2009 Aug;38(1):1-3. [doi: 10.1007/s12160-009-9132-5] [Medline: 19816750]

79.   Ritterband LM, Thorndike FP, Cox DJ, Kovatchev BP, Gonder-Frederick LA. A behavior change model for internet interventions. Ann Behav Med 2009 Aug;38(1):18-27 [FREE Full text] [doi: 10.1007/s12160-009-9133-4] [Medline: 19802647]

80.   Tate DF, Finkelstein EA, Khavjou O, Gustafson A. Cost effectiveness of internet interventions: review and recommendations. Ann Behav Med 2009 Aug;38(1):40-45 [FREE Full text] [doi: 10.1007/s12160-009-9131-6] [Medline: 19834778]

81.   Rogers E. Diffusion of Innovations. New York: Free Press; 2008.

XSL•FO
RenderX

82. Christensen H, Mackinnon A. The law of attrition revisited. J Med Internet Res 2006;8(3):e20; author reply e21 [FREE Full text] [doi: 10.2196/jmir.8.3.e20] [Medline: 17032636]

83. Wangberg SC, Bergmo TS, Johnsen JA. Adherence in Internet-based interventions. Patient Prefer Adherence 2008;2:57-65 [FREE Full text] [Medline: 19920945]

84. Glasgow RE. Enhancing the scientific foundation of internet intervention research. Ann Behav Med 2009 Aug;38(1):46-47. [doi: 10.1007/s12160-009-9134-3] [Medline: 19816749]

85. Napolitano MA, Fotheringham M, Tate D, Sciamanna C, Leslie E, Owen N, et al. Evaluation of an internet-based physical activity intervention: a preliminary investigation. Ann Behav Med 2003;25(2):92-99. [Medline: 12704010]

86. Koo M, Skinner H. Challenges of internet recruitment: a case study with disappointing results. J Med Internet Res 2005;7(1):e6 [FREE Full text] [doi: 10.2196/jmir.7.1.e6] [Medline: 15829478]

87. Lieberman DZ, Massey SH. Pathways to change: the effect of a Web application on treatment interest. Am J Addict 2008;17(4):265-270. [doi: 10.1080/10550490802138525] [Medline: 18612880]

88. Albarracín D, Durantini MR, Earl A, Gunnoe JB, Leeper J. Beyond the most willing audiences: a meta-intervention to increase exposure to HIV-prevention programs by vulnerable populations. Health Psychol 2008 Sep;27(5):638-644. [doi: 10.1037/0278-6133.27.5.638] [Medline: 18823190]

89. Brouwer W, Oenema A, Crutzen R, De Nooijer J, De Vries NK, Brug J. What makes people decide to visit and use an internet-delivered behavior-change intervention?: A qualitative study among adults. Health Education 2009;109:460-473.

90. Crutzen R, Cyr D, de Vries NK. Bringing loyalty to e-Health: theory validation using three internet-delivered interventions. J Med Internet Res 2011;13(3):e73 [FREE Full text] [doi: 10.2196/jmir.1837] [Medline: 21946128]

91. Crutzen R, Peters GJ, Abraham C. What about trialists sharing other study materials? BMJ 2012;345:e8352. [Medline: 23229066]

92. Peters G-JY, Abraham C, Crutzen R. Eur Health Psychol. 2012. Full disclosure: doing behavioural science necessitates sharing URL: http://www.ehps.net/ehp/issues/2012/v14iss4_December2012/14_4_Peters,Abraham,Crutzen.pdf [accessed 2013-07-11] [WebCite Cache ID 6I1tGHMZp]

93. Burda BU, Norris SL, Holmer HK, Ogden LA, Smith ME. Quality varies across clinical practice guidelines for mammography screening in women aged 40-49 years as assessed by AGREE and AMSTAR instruments. J Clin Epidemiol 2011 Sep;64(9):968-976. [doi: 10.1016/j.jclinepi.2010.12.005] [Medline: 21420280]

Original Paper

Efficacy of Standard Versus Enhanced Features in a Web-Based Commercial Weight-Loss Program for Obese Adults, Part 2: Randomized Controlled Trial

Clare E Collins[1], PhD; Philip J Morgan[2], PhD; Melinda J Hutchesson[1], PhD; Robin Callister[3], PhD

[1]Priority Research Centre in Nutrition and Physical Activity, Nutrition and Dietetics, School of Health Sciences, Faculty of Health, University of Newcastle, Callaghan, Australia

[2]Priority Research Centre in Nutrition and Physical Activity, School of Education, Faculty of Education & Arts, University of Newcastle, Callaghan, Australia

[3]Priority Research Centre in Nutrition and Physical Activity, School of Biomedical Sciences & Pharmacy, Faculty of Health, University of Newcastle, Callaghan, Australia

Corresponding Author:
Clare E Collins, PhD
Priority Research Centre in Nutrition and Physical Activity, Nutrition and Dietetics
School of Health Sciences, Faculty of Health
University of Newcastle
Level 2, ATC Building
University Drive
Callaghan, 2308
Australia
Phone: 61 2 4921 5646
Fax: 61 2 4921 7053
Email: Clare.Collins@newcastle.edu.au

Abstract

Background: Commercial Web-based weight-loss programs are becoming more popular and increasingly refined through the addition of enhanced features, yet few randomized controlled trials (RCTs) have independently and rigorously evaluated the efficacy of these commercial programs or additional features.

Objective: To determine whether overweight and obese adults randomized to an online weight-loss program with additional support features (enhanced) experienced a greater reduction in body mass index (BMI) and increased usage of program features after 12 and 24 weeks compared to those randomized to a standard online version (basic).

Methods: An assessor-blinded RCT comparing 301 adults (male: n=125, 41.5%; mean age: 41.9 years, SD 10.2; mean BMI: 32.2 kg/m^2, SD 3.9) who were recruited and enrolled offline, and randomly allocated to basic or enhanced versions of a commercially available Web-based weight-loss program for 24 weeks.

Results: Retention at 24 weeks was greater in the enhanced group versus the basic group (basic 68.5%, enhanced 81.0%; P=.01). In the intention-to-treat analysis of covariance with imputation using last observation carried forward, after 24 weeks both intervention groups had reductions in key outcomes with no difference between groups: BMI (basic mean –1.1 kg/m^2, SD 1.5; enhanced mean –1.3 kg/m^2, SD 2.0; P=.29), weight (basic mean –3.3 kg, SD 4.7; enhanced mean –4.0 kg, SD 6.2; P=.27), waist circumference (basic mean –3.1 cm, SD 4.6; enhanced mean –4.0 cm, SD 6.2; P=.15), and waist-to-height ratio (basic mean –0.02, SD 0.03; enhanced mean –0.02, SD 0.04, P=.21). The enhanced group logged in more often at both 12 and 24 weeks, respectively (enhanced 12-week mean 34.1, SD 28.1 and 24-week mean 43.1, SD 34.0 vs basic 12-week mean 24.6, SD 25.5 and 24-week mean 31.8, SD 33.9; P=.002).

Conclusions: The addition of personalized e-feedback in the enhanced program provided limited additional benefits compared to a standard commercial Web-based weight-loss program. However, it does support greater retention in the program and greater usage, which was related to weight loss. Further research is required to develop and examine Web-based features that may enhance engagement and outcomes and identify optimal usage patterns to enhance weight loss using Web-based programs.

XSL•FO
RenderX

Introduction

Internationally, obesity rates in adults continue to rise unabated [1]. Effective treatment programs with broad reach are urgently required. Web-based weight-loss programs are an increasingly viable option because most US and Australian households (66% [2] and 72% [3], respectively) have access to broadband Internet, and many adults (61% in the United States) seek information on health, nutrition, and weight loss from the Internet [4].

A systematic review of the effectiveness of Web-based weight loss and maintenance interventions found that these programs can facilitate meaningful weight change [5]. However, it was not possible to determine their overall effectiveness because of the heterogeneity of designs and small number of comparable studies. A meta-analysis of 3 Web-based weight-loss randomized controlled trials (RCTs) that compared online education-only programs with online programs that included enhanced features, such as counseling, automated or therapist feedback, behavioral lessons, self-monitoring, and a bulletin board, found weight loss was increased by 2.2 kg over a 6- to 12-month period [5]. These results are supported by 3 other RCTs which found that the addition of online lessons with daily self-monitoring of weight, eating, and exercise and computer-generated feedback [6], or the addition of peer support [7], or individually tailored action plans [8], resulted in greater weight loss after 24 weeks [6], a trend toward a greater effect size after 12 weeks [7] and greater weight loss [8] compared to an online program without the enhanced features. Krukowski et al [9] have also demonstrated that participant's usage of feedback components of a Web-based weight-loss program (eg, progress charts) was the most significant predictor of weight loss after 6 months. By contrast, 2 other RCTs found that adding online lessons or a weekly online group chat session to a Web-based weight-loss program was equally effective up to 12 and 16 weeks, respectively, as a Web-based program without these features [6,10]. Further, all these studies were conducted in the United States [6-10]. Additional longer-term studies from other regions of the world are required to evaluate the superiority, or otherwise, of Web-based programs with enhanced features.

Within the currently available online commercial weight-loss programs, there is a large degree of variation across the range of features provided, including blogs, chat rooms, self-monitoring tools for weight, diet, and physical activity, and also differing types and amounts of feedback from generic to tailored information and human e-counseling. To date, the ability of these more personalized enhanced features to facilitate greater weight loss has only had limited evaluation because programs have not tracked use of specific features [11].

We have previously compared the efficacy of a standard commercial Web-based weight-loss program (basic) versus an enhanced version of this Web program that provided additional personalized e-feedback and contact from the provider (enhanced) versus a waitlist control group [12,13]. After 12 weeks, we found both Web-based programs produced significantly greater weight loss and reductions in body mass index (BMI) compared to the waiting list control group, but no differences in the weight-related outcomes were observed between the 2 programs. Part 2 of the study aims to determine whether overweight and obese adults randomized to the enhanced version of the commercial Web-based weight-loss program achieve a larger reduction in BMI and usage of program features compared to those randomized to a standard version of the online program without these features after 24 weeks.

Methods

This assessor-blinded RCT recruited overweight and obese adults from the Hunter community in New South Wales, Australia, who were enrolled offline in 2009. Eligibility criteria included age 18 to 60 years, BMI 25 to 40 kg/m^2, not participating in other weight-loss programs, pass a health screen [14], available for in-person assessments, and access to a computer with email and Internet services. Written informed consent was obtained from all participants, and ethical approval obtained from the University of Newcastle Human Ethics Research Committee. The trial conformed to the Consolidated Standards of Reporting Trials (CONSORT)-eHealth Checklist (Multimedia Appendix 1) [15].

Stratification and Randomization

After baseline assessments were completed, participants were stratified by sex and BMI category (25 to <30, ≥30 to <35, or ≥35 to 40 kg/m^2) and randomized using a stratified block design to either the standard (basic) Web-based weight-loss program or the same program with additional features (enhanced) (Figure 1). At baseline, participants could also have been randomized to a waitlist control group who were not provided with access to the weight-loss program website. After 12 weeks, participants in the control group were rerandomized into either the basic or enhanced groups and data collected after this rerandomization were included in this analysis. Participants who dropped out before rerandomization to a treatment arm, or achieved their weight-loss goal (≥10% of baseline weight lost) and had, therefore, entered the weight maintenance phase, were also excluded from this analysis.

Figure 1. Participant flow.
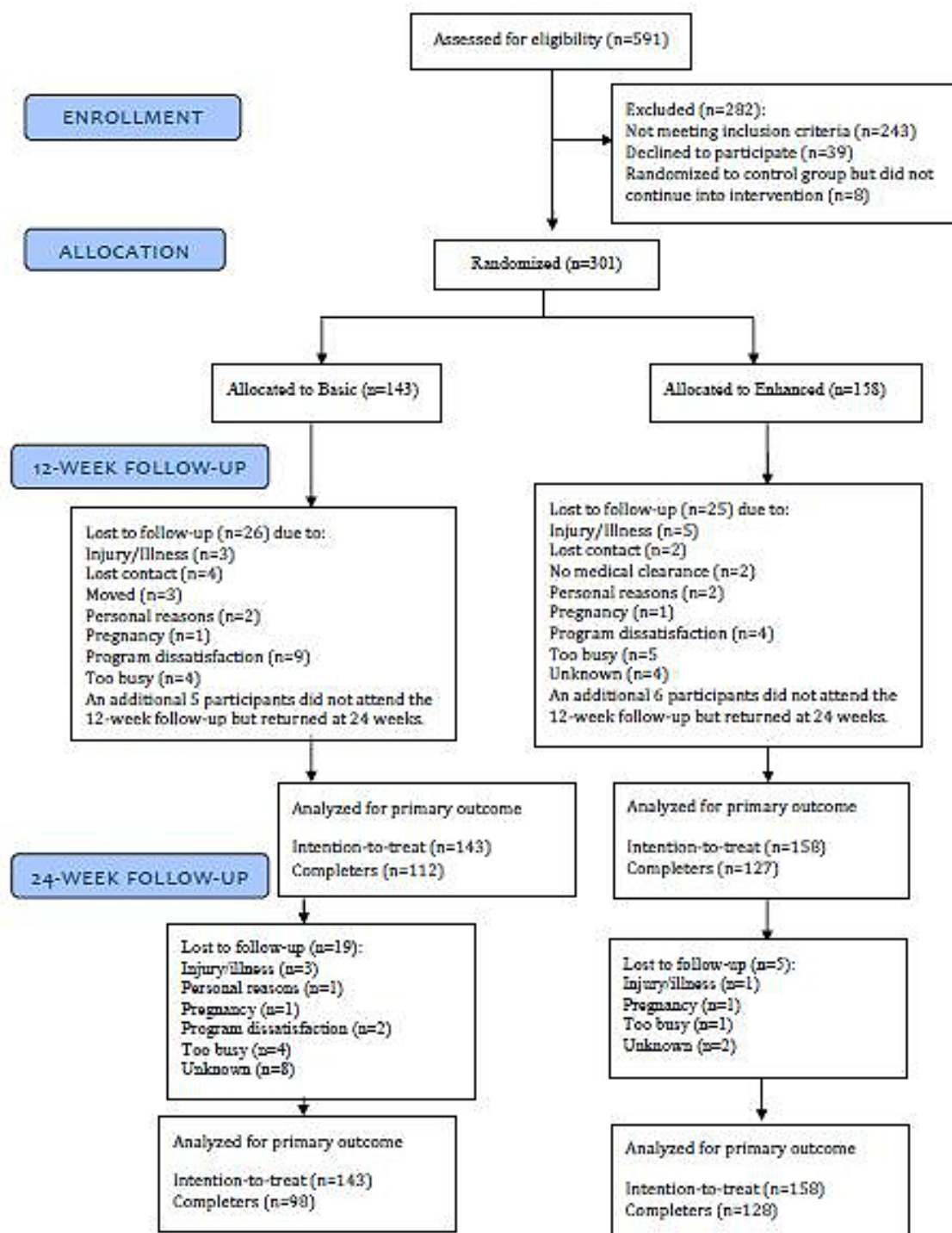
Web-Based Weight-Loss Programs (Basic and Enhanced)

Participants were provided with free access to the basic or enhanced version of a commercial Web-based program provided by SP Health Co Pty Ltd in Australia under the name The Biggest Loser Club. The basic program was the version commercially available at the time of the study (2009-2010).

Program features are reported in Table 1. The enhanced program contained additional features to the basic program and was provided in a closed test environment. At baseline, participants were given instructions to log in and set up their program details. They were also given a company phone number in case they experienced any difficulties in logging in. Participants did not receive any training on program use to mirror the commercial program engagement experience and increase external validity.

Participants were blinded to group allocation and accessed the website using their usual Internet connection.

The 12-week Web-based programs were based on social cognitive theory [17]. Key behavior change mediators targeted included self-efficacy, goal setting, self-monitoring, outcome expectations, and social support. An individualized daily energy intake target to facilitate a weight loss of 0.5 to 1 kg per week was set, as well as a goal weight. Participants were encouraged to self-monitor by reporting their weight or other body measurements via the website or short message service (SMS) text messages once per week and could view graphs and charts to track their progress overtime. They were also encouraged to self-monitor their dietary intake and exercise using an online diary at least 4 days per week. The diary provided automated feedback on daily and weekly energy intake and expenditure, and a weekly summary macronutrient and micronutrient intake compared with recommended targets. Social support was available via a discussion board. Online information was provided weekly (calorie-controlled, low-fat menu plans and grocery lists; physical activity plan based on exercise preferences; educational tips and challenges) which participants were prompted to access via a weekly email newsletter.

At the end of 12 weeks, participants could choose to repeat the same weekly 12-week program or to continue for an additional 12 weeks with content varied based on the season and keeping their entire accumulated personal progress and data.

The enhanced program included all the basic program features described previously. The additional components were: (1) personalized, system-generated enrollment reports that suggested appropriate weight-loss goals and key behavior changes required for success based on response to a behavioral survey at enrollment; (2) weekly automated system-generated, personalized e-feedback for key elements of diet and physical activity based on diary entries, usage patterns of website features, and level of success with weight loss (Figure 2); (3) an escalating reminder schedule to use the diary, visit the program site, and enter a weekly weight (an initial reminder email, then a SMS text message if there was no response, then a reminder phone call if a weekly weight was still not entered).

Table 1. Description of the basic and enhanced commercial Web-based weight-loss programs.

Basic and enhanced	Enhanced only
Participants set weight-loss goals, advised to self-monitor their weight, waist, and hip girths. Encouraged to self-monitor via weekly email and/or short message service (SMS) text messaging reminders to enter weight on website. Entered data were tracked and displayed graphically and in a body (BMI) silhouette.	Personalized automated enrollment reports suggesting appropriate weight-loss goals and key behavior changes required for success. Eating behaviors targeted included total energy, saturated fat and fiber intake, daily servings of fruit and vegetables, high-risk eating behaviors (eg, skipping meals, not eating breakfast, drinking soft drinks) and nonhungry eating triggers.
Individualized daily calorie targets to facilitate 0.5-1 kg weight loss per week (~2600 kJ less than their estimated energy requirements).	Weekly automated personalized feedback for key elements of diet and physical activity based on diary entries; usage patterns for website features; and level of success with weight loss. Eating behaviors targeted were consistent with the enrollment reports (Figure 2).
Access to weekly low-fat menu plan and grocery lists designed to meet nutrient reference values [16] and assigned calorie target.	Reminders to use the online diary, visit the site, and/or weigh-in. The reminder schedule included an initial reminder email; if no response, a text message; if no response, a phone call.
Web-based food and exercise diary to monitor energy intake and energy expenditure. Daily and weekly calculations of energy balance and nutrition summaries compared with recommended nutrient targets if food entries made in online diary.	
Online education in the form of weekly tutorials, fact sheets, meal, and exercise plans and weekly challenges.	
Social support via online discussion forums.	

Outcome Measures

Participant assessments were conducted at the University of Newcastle at baseline, 12, and 24 weeks. Blinded research assistants conducted assessments for all groups, and participants were reminded at each assessment not to discuss group allocation.

Height was measured to 0.1 cm using the stretch stature method on a Harpenden portable stadiometer (Holtain Limited, Dyfed, UK). Weight was measured in light clothing, without shoes on a digital scale to 0.01 kg (CH-150kp, A&D Mercury Pty Ltd, Australia) and the primary outcome of BMI (kg/m^2) calculated as weight (kg)/height (m)^2. Waist circumference was measured to 0.1 cm using a nonextensible steel tape (KDSF10-02, KDS

Corporation, Osaka, Japan) at 2 points: (1) level with the umbilicus and (2) at the narrowest point between the lower costal border and the umbilicus. Waist-to-height ratio was then calculated. Blood pressure and heart rate were measured using an automated blood pressure monitor (NISSEI/DS-105E digital electronic blood pressure monitor; Nihon Seimitsu Sokki Co Ltd, Gunma, Japan) under standardized conditions. Blood samples were collected with participants advised to fast overnight and analyzed for lipids (total cholesterol, low-density lipoprotein [LDL] and high-density lipoprotein [HDL], cholesterol, and triglycerides), glucose, and insulin using standard automated techniques at a single National Association of Testing Authorities accredited pathology service.

Dietary intake was assessed using the Australian Eating Survey (AES), a 120-item semiquantitative food-frequency questionnaire (FFQ). The AES has been evaluated for reliability and relative validity and demonstrates acceptable accuracy for ranking nutrient intakes in Australian adults [18]. Nutrient intakes are calculated using the Australian food composition database [19], and analyzed using a standard protocol.

The 18-item Three-Factor Eating Questionnaire-R18 (TFEQ-R18) was used to measure cognitive restraint, uncontrolled eating, and emotional eating [20]. Quality of life was assessed using the SF-36 version 2.0 (QualityMetric Incorporated, Lincoln, RI, USA), a multipurpose, generic short-form health survey consisting of an 8-scale profile of functional health and well-being scores and psychometrically based physical and mental health summary measures [21].

The International Physical Activity Questionnaire-short form (IPAQ-SF) was used to estimate total metabolic equivalent (MET)-minutes/week [22]. Pedometers were used to measure steps per day for 7 consecutive days (Yamax SW700; Yamax Corporation, Kumamoto City, Japan) with step counts adjusted for additional self-reported physical activity (eg, contact sports, swimming, cycling).

Statistical Analysis

Continuous data were summarized using descriptive statistics including mean (SD) and categorical data as category percentages. Demographic and baseline variables were compared between treatment groups using analysis of variance (ANOVA) for continuous variables and chi-square tests for categorical variables. Analysis of covariance (ANCOVA) was used to test for differences in outcomes at 12 weeks and 24 weeks between treatment groups after adjusting for the baseline value of that outcome. The model outcome was the variable of interest at 12 or 24 weeks with the baseline level used as a covariate. The only other variable included in the model was sex. Differences and 95% confidence intervals between treatment groups in the outcome at each time point were estimated using the least squares means from the ANCOVA models.

The intention-to-treat (ITT) population includes all participants who were randomized into 1 of the 2 treatment groups. For participants who had missing data at 12 or 24 weeks, their missing data was imputed using the last observation carried forward (LOCF) and baseline observation carried forward (BOCF) approach. The completer population includes all individuals who attended the 24-week assessment, and subgroup analyses are based on this population.

An additional analysis was conducted using a generalized linear mixed model (GLMM) to test for a difference between groups across the combined 12-week and 24-week time points. The outcome in this model was the individual's outcome at the 2 posttreatment assessments; the main predictor of interest was treatment group with the baseline value of the outcome included as a covariate. Sex was also included as a covariate in these models because it is a common confounding factor. All analyses were programmed in Stata v11 or SAS v9.2 (StataCorp LP, College Station, TX, USA).

Figure 2. Enhanced groups weekly automated personalized feedback.
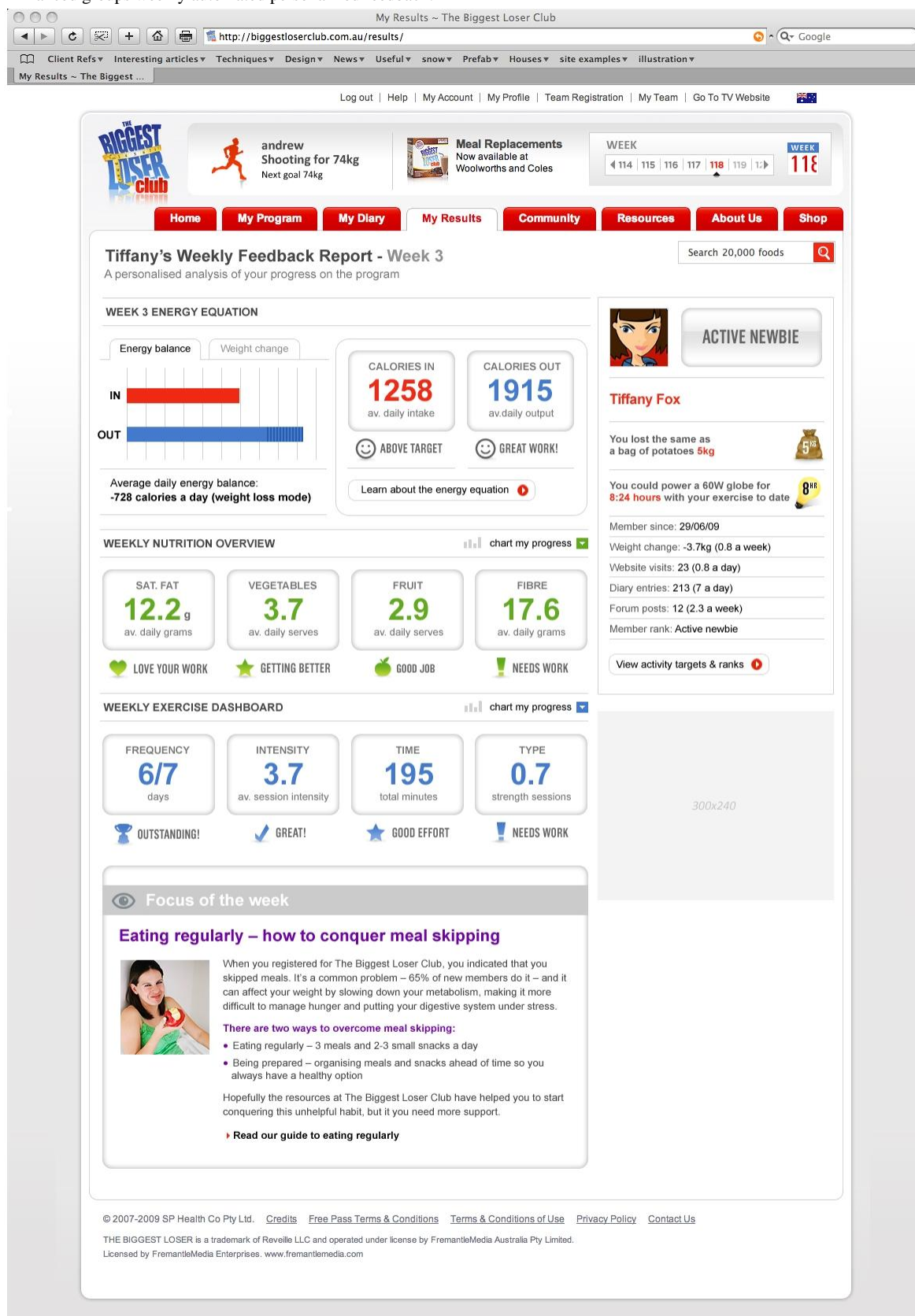
Results

Baseline Characteristics

Of the 591 participants assessed for eligibility, 309 (129 males, 180 females) were initially randomized into the 3 groups (basic: n=99, enhanced: n=106, or waitlist control: n=104). After 12 weeks, the control group, of whom 8 were lost to follow-up, were rerandomized (96 participants, 52 enhanced, 44 basic) into the trial. Therefore, in the current analysis, 301 participants (125 males, 176 females) were randomized to the basic (n=143) or enhanced (n=158) groups (Figure 1).

Participants who were randomized to the basic group were similar at baseline to those randomized to the enhanced group for all demographic and other baseline characteristics (Table 2). Mean age of participants was 42 years (SD 10.2), most were overweight or obese (BMI 30 to <35), Australian born, and reported a weekly household income of ≥AU $1500.

Retention at 12 and 24 Weeks

Participant flow through the trial (Figure 1) shows the number of participants who were randomized to each treatment condition, the number who withdrew with reasons, and the number who had data at 12 and 24 weeks. There was no significant difference in retention rates between the basic (74.7%) and enhanced (84.9%) groups after 12 weeks (*P*=.66); however, more enhanced group participants attended the 24-week assessments (basic 68.5%, enhanced 81.0%, *P*=.01).

Changes in Weight, Body Mass Index, and Waist Circumference

Weight, BMI, and waist circumference were significantly lower than baseline at 12 and 24 weeks in each group. Change in the primary outcome of BMI was similar between those randomized to the basic and enhanced groups at 12 and 24 weeks after treatment in the LOCF (Table 3), BOCF (Table 4), and completers (Table 5) analyses. For the LOCF (basic mean -3.6, SD 4.9; enhanced mean -4.3, SD 6.4), BOCF (basic mean -3.2, SD 4.7; enhanced mean -4.2, 6.3), and completers analysis (basic mean -3.9, SD 4.1; enhanced mean -4.6, SD 4.8), there

were no significant between-group differences for the mean percentage weight loss at 24 weeks or the proportion of participants achieving clinically important weight losses of ≥5% [23] at 24 weeks (LOCF: basic 31.5%, enhanced 38.0%; BOCF: basic 28.7%, enhanced 36.7%; completers: basic 41.2%, enhanced 45.7%).

Secondary Outcomes

There was only 1 significant difference in secondary outcomes between the basic and enhanced groups in the LOCF (Table 3), BOCF (Table 4), and completers (Table 5) analyses. The BOCF analyses found that the enhanced group demonstrated a significantly greater (*P*=.03) reduction in resting heart rate than the basic group after 24 weeks.

Subgroup Analyses

The change in primary and secondary outcomes within treatment groups was similar across all subgroups (sex, age, BMI category) of the completer population at 12 or 24 weeks (data not presented). There were no statistically significant interactions between treatment group and sex (*P*=.52), treatment group and BMI category (*P*=.45), or treatment group and age group (*P*=.72) for the outcome of weight.

Website Usage

There was a significantly greater website usage in the enhanced group compared to the basic group at both 12 and 24 weeks with the enhanced group logging on an additional 10 days over the first 12 weeks and 12 days over 24 weeks (*P*=.002) (Table 6). A similar result was found for the completers population (*P*=.02).

In the completers population, significant correlations were found between the percentage weight loss at 12 and 24 weeks and total website usage from baseline to 12 weeks (*r* = –0.50, *P*<.001) and 24 weeks (*r* = –0.50, *P*<.001), respectively (data not shown). Participants who achieved clinically significant (≥5%) weight loss at 12 and 24 weeks used the website on significantly more days from baseline to 12 weeks (median 44 vs 13 days, *P*<.001) and 24 weeks (median 58 vs 16 days, *P*<.001) than those with <5% weight loss.

Table 2. Demographic and other baseline characteristics by treatment group.

Characteristic	Treatment group		Total (N=301)	P value[a]
	Basic (n=143)	Enhanced (n=158)		
Sex (male), n (%)	59 (47.2)	66 (52.8)	125 (41.5)	.93
BMI group strata				
25 to <30	50 (46.7)	57 (53.3)	107 (35.5)	.97
30 to <35	58 (48.3)	62 (51.7)	120 (39.9)	
35 to <40	35 (47.3)	39 (52.7)	74 (24.6)	
Current or previous smoker (never smoked), n (%)	128 (47.8)	140 (52.2)	268 (89.0)	.91
Highest level of education, n (%)				
High school	43 (47.8)	47 (52.2)	90 (29.90)	.60
Trade/diploma	46 (44.2)	58 (55.8)	104 (34.6)	
University degree	37 (54.4)	31 (45.60)	68 (2.6)	
Higher university degree	17 (44.7)	21 (55.3)	38 (12.6)	
Weekly household income (AU $), n (%)				
<$700	12 (50.0)	12 (50.0)	24 (8.0)	.77
$700 to <$1000	8 (50.0)	8 (50.0)	16 (5.6)	
$1000 to <$1400	20 (57.1)	15 (42.9)	35 (11.6)	
≥$1500	92 (46.0)	108 (54.0)	200 (66.4)	
Country of birth (Australia), n (%)	132 (48.4)	141 (51.6)	273 (90.7)	.45
Age (years), mean (SD)	41.9 (10.1)	42.0 (10.3)	41.9 (10.2)	.91
Height (m), mean (SD)	1.7 (0.1)	1.7 (0.1)	1.7 (0.1)	.53
Weight (kg), mean (SD)	94.4 (15.5)	93.4 (13.9)	93.9 (14.7)	.53
Body mass index (kg/m^2), mean (SD)	32.2 (3.7)	32.2 (4.1)	32.2 (3.9)	.95
Waist circumference at umbilicus (cm), mean (SD)	106.8 (10.2)	106.7 (11.5)	106.7 (10.9)	.98
Waist circumference at narrowest point (cm), mean (SD)	98.3 (11.6)	97.8 (11.2)	98.1 (11.0)	.70
Waist-to-height ratio at umbilicus, mean (SD)	0.6 (0.1)	0.6 (0.1)	0.6 (0.1)	.76
Waist-to-height ratio at narrowest point, mean (SD)	0.58 (0.06)	0.58 (0.06)	0.58 (0.06)	.94
Systolic blood pressure (mmHg), mean (SD)	121 (13)	121 (12)	121 (12)	.97
Diastolic blood pressure (mmHg), mean (SD)	79 (10)	79 (10)	79 (10)	.88
Resting heart rate (bpm), mean (SD)	69 (11)	68 (10)	68 (10)	.56
Total serum cholesterol (mmol/L), mean (SD)	5.1 (0.9)	5.1 (1.0)	5.1 (1.0)	.54
LDL cholesterol (mmol/L), mean (SD)	3.1 (0.8)	3.0 (0.9)	3.1 (0.8)	.26
HDL cholesterol (mmol/L), mean (SD)	1.3 (0.3)	1.3 (0.3)	1.3 (0.3)	.78
Triglycerides (mmol/L), mean (SD)	1.6 (0.8)	1.8 (1.1)	1.7 (1.0)	.09
LDL to HDL ratio, mean (SD)	2.53 (0.79)	2.39 (0.77)	2.46 (0.78)	.14
Glucose (mmol/L), mean (SD)	4.7 (0.7)	4.7 (0.6)	4.7 (0.7)	.35
Insulin (mIU/L), mean (SD)	11.1 (11.7)	10.5 (11.3)	10.8 (11.4)	.67
Physical functioning (SF36), mean (SD)	85.9 (13.5)	83.9 (18.4)	84.9 (16.3)	.30
Mental health (SF36), mean (SD)	73.5 (16.1)	73.2 (17.4)	73.4 (16.8)	.89
Total physical activity MET (min/week), mean (SD)	3028 (3188)	2877 (3100)	2948 (3137)	.69
Cognitive restraint scale, mean (SD)	13.3 (3.0)	13.4 (2.9)	13.3 (3.0)	.70

XSL•FO
RenderX

Characteristic	Treatment group		Total (N=301)	P value[a]
	Basic (n=143)	Enhanced (n=158)		
Uncontrolled eating scale, mean (SD)	21.1 (4.8)	20.9 (5.3)	21.0 (5.0)	.63
Emotional eating score, mean (SD)	7.7 (2.3)	7.7 (2.6)	7.7 (2.5)	.93
Total energy intake, mean (SD)	9983 (3278)	9972 (3236)	9977 (3251)	.98

[a]P values are from ANOVA for continuous measures and from a chi-square tests for categorical measures.

Table 3. Mean change in a range of variables from baseline to 12 weeks and baseline to 24 weeks within each treatment group and the least squares mean (LSM) difference in change between treatment groups (ITT population LOCF approach).

Characteristic and follow-up time	Treatment group Mean change (SD)		Absolute difference between groups LSM (95% CI)	P values for group effect	
	Basic	Enhanced	Enhanced vs basic	Difference at follow-up	Difference between groups
Weight (kg)					
12 weeks	–2.7 (4.0)	–3.3 (4.5)	0.6 (–0.3, 1.6)	.21	
24 weeks	–3.3 (4.7)	–4.0 (6.2)	0.7 (–0.6, 2.0)	.27	.21
Percentage weight loss (%)					
12 weeks	–2.90 (4.09)	–3.61 (4.69)	0.71 (–0.30, 1.71)	.17	
24 weeks	–3.56 (4.94)	–4.28 (6.38)	0.71 (–0.59, 2.02)	.28	.21
Attained 5% weight loss (%)					
12 weeks	24.5 (43.1)	32.9 (47.1)	8.4 (–1.9, 18.7)	.11	
24 weeks	31.5 (46.6)	38.0 (48.7)	6.5 (–4.3, 17.4)	.23	.13
Systolic blood pressure (mmHg)					
12 weeks	–3.40 (10.31)	–3.94 (9.66)	0.52 (–1.60, 2.65)	.63	
24 weeks	–3.00 (10.11)	–2.33 (11.20)	0.70 (–1.46, 2.86)	.53	.93
Diastolic blood pressure (mmHg)					
12 weeks	–1.82 (8.50)	–2.30 (7.57)	0.54 (–1.14, 2.22)	.53	
24 weeks	–1.18 (8.97)	–1.03 (8.05)	0.09 (–1.68, 1.86)	.92	.78
Body mass index (kg/m^2)					
12 weeks	–0.9 (1.3)	–1.1 (1.5)	0.2 (–0.1, 0.5)	.28	
24 weeks	–1.1 (1.5)	–1.3 (2.0)	0.2 (–0.2, 0.6)	.29	.27
Resting heart rate (bpm)					
12 weeks	–1.33 (7.12)	–2.35 (6.49)	1.20 (–0.23, 2.64)	.10	
24 weeks	–1.62 (7.69)	–3.03 (7.22)	1.59 (–0.02, 3.19)	.05	.05
Waist circumference at umbilicus (cm)					
12 weeks	–3.4 (4.5)	–3.6 (5.3)	0.2 (–0.9, 1.3)	.73	
24 weeks	–4.4 (5.3)	–5.3 (6.7)	0.9 (–0.5, 2.3)	.22	.38
Waist circumference at narrowest point (cm)					
12 weeks	–2.5 (4.3)	–3.4 (4.8)	0.9 (–0.1, 1.9)	.09	
24 weeks	–3.1 (4.6)	–4.0 (6.2)	0.9 (–0.3, 2.2)	.15	.10
Waist-to-height ratio at umbilicus					
12 weeks	–0.02 (0.03)	–0.02 (0.03)	0.00 (–0.01, 0.01)	.86	
24 weeks	–0.03 (0.03)	–0.03 (0.04)	0.00 (–0.00, 0.01)	.28	.49
Waist-to-height ratio at narrowest point					
12 weeks	–0.01 (0.02)	–0.02 (0.03)	0.00 (–0.00, 0.01)	.11	
24 weeks	–0.02 (0.03)	–0.02 (0.04)	0.00 (–0.00, 0.01)	.21	.14
Total serum cholesterol (mmol/L)					
12 weeks	–0.08 (0.50)	–0.19 (0.58)	0.12 (–0.00, 0.24)	.06	
24 weeks	0.01 (0.52)	–0.08 (0.62)	0.09 (–0.03, 0.22)	.15	.06

XSL•FO
RenderX

Characteristic and follow-up time	Treatment group Mean change (SD)		Absolute difference between groups LSM (95% CI)	*P* values for group effect	
	Basic	Enhanced	Enhanced vs basic	Difference at follow-up	Difference between groups
LDL cholesterol (mmol/L)					
12 weeks	–0.00 (0.45)	–0.07 (0.46)	0.09 (–0.02, 0.19)	.11	
24 weeks	0.04 (0.49)	–0.01 (0.50)	0.07 (–0.05, 0.18)	.27	.13
HDL cholesterol (mmol/L)					
12 weeks	–0.00 (0.14)	–0.01 (0.16)	0.00 (–0.03, 0.04)	.93	
24 weeks	0.02 (0.13)	0.02 (0.17)	0.00 (–0.03, 0.04)	.80	.92
Triglycerides (mmol/L)					
12 weeks	–0.16 (0.63)	–0.26 (0.66)	0.05 (–0.08, 0.18)	.48	
24 weeks	–0.09 (0.64)	–0.23 (0.66)	0.09 (–0.04, 0.23)	.18	.30
LDL to HDL ratio					
12 weeks	0.00 (0.40)	–0.04 (0.35)	0.06 (–0.03, 0.15)	.21	
24 weeks	–0.01 (0.42)	–0.04 (0.37)	0.04 (–0.05, 0.14)	.35	.21
Glucose (mmol/L)					
12 weeks	–0.22 (0.55)	–0.22 (0.55)	0.02 (–0.10, 0.14)	.78	
24 weeks	–0.07 (0.46)	–0.12 (0.54)	0.08 (–0.03, 0.18)	.17	.37
Insulin (mIU/L)					
12 weeks	–1.76 (10.91)	–1.80 (5.47)	0.29 (–1.25, 1.82)	.71	
24 weeks	–2.51 (10.36)	–2.31 (6.18)	0.04 (–1.47, 1.55)	.96	.81
Physical functioning (SF36)					
12 weeks	2.48 (20.05)	4.71 (15.74)	1.14 (–2.34, 4.62)	.52	
24 weeks	3.58 (10.89)	4.62 (15.03)	0.17 (–2.30, 2.64)	.89	.63
Mental health (SF36)					
12 weeks	2.13 (15.06)	4.04 (13.43)	1.58 (–1.41, 4.57)	.30	
24 weeks	1.77 (14.83)	3.82 (23.80)	1.68 (–2.69, 6.04)	.45	.31
Total physical activity MET (min/week)					
12 weeks	215.29 (2448.3)	373.56 (2467.1)	96.19 (–424.3, 616.66)	.72	
24 weeks	203.58 (2668.5)	619.84 (3156.7)	358.58 (–285.1, 1002.3)	.27	.36
Average step count per day					
12 weeks	319.92 (2481.3)	1059.8 (3094.2)	675.22 (–11.32, 1361.8)	.05	
24 weeks	94.15 (2303.2)	706.99 (3173.9)	548.38 (–142.9, 1239.6)	.12	.06
Cognitive restraint scale					
12 weeks	1.32 (2.72)	1.74 (3.28)	0.48 (–0.15, 1.11)	.14	
24 weeks	1.36 (2.96)	1.90 (3.37)	0.59 (–0.09, 1.27)	.09	.08
Uncontrolled eating scale					
12 weeks	–1.71 (3.44)	–1.78 (3.41)	0.14 (–0.58, 0.86)	.70	
24 weeks	–1.67 (3.59)	–1.76 (3.62)	0.17 (–0.59, 0.93)	.66	.65
Emotional eating score					
12 weeks	–0.37 (1.38)	–0.48 (1.56)	0.10 (–0.22, 0.41)	.54	
24 weeks	–0.45 (1.68)	–0.61 (1.72)	0.15 (–0.21, 0.51)	.41	.43

XSL•FO
RenderX

Characteristic and follow-up time	Treatment group Mean change (SD)		Absolute difference between groups LSM (95% CI)	P values for group effect	
	Basic	Enhanced	Enhanced vs basic	Difference at follow-up	Difference between groups
Total energy intake					
12 weeks	–985.5 (2386.7)	–1248 (2458.1)	267.74 (–186.4, 721.90)	.25	
24 weeks	–952.4 (2293.8)	–929.7 (2363.6)	18.23 (–439.2, 475.66)	.94	.57

Table 4. Mean change in a range of variables from baseline to 12 weeks and baseline to 24 weeks within each treatment group and the least squares mean (LSM) difference in change between treatment groups (ITT population BOCF approach).

Characteristic and follow-up time	Treatment group Mean change (SD)		Absolute difference between groups LSM (95% CI)	P values for group effect	
	Basic	Enhanced	Enhanced vs basic	Difference at follow-up	Difference between groups
Weight (kg)					
12 weeks	–2.7 (4.0)	–3.3 (4.5)	0.6 (–0.3, 1.6)	.21	
24 weeks	–3.0 (4.5)	–3.9 (6.2)	1.0 (–0.2, 2.2)	.11	.13
Percentage weight loss (%)					
12 weeks	–2.90 (4.09)	–3.61 (4.69)	0.71 (–0.30, 1.71)	.17	
24 weeks	–3.17 (4.74)	–4.19 (6.34)	1.02 (–0.27, 2.30)	.12	.12
Attained 5% weight loss (%)					
12 weeks	24.5 (43.1)	32.9 (47.1)	8.43 (–1.87, 18.73)	.11	
24 weeks	28.7 (45.4)	36.7 (48.4)	8.07 (–2.58, 18.73)	.13	.09
Systolic blood pressure (mmHg)					
12 weeks	–3.40 (10.31)	–3.94 (9.66)	0.52 (–1.60, 2.65)	.63	
24 weeks	–2.46 (9.71)	–2.14 (10.85)	0.35 (–1.77, 2.47)	.75	.92
Diastolic blood pressure (mmHg)					
12 weeks	–1.82 (8.50)	–2.30 (7.57)	0.54 (–1.14, 2.22)	.53	
24 weeks	–0.89 (8.59)	–0.91 (7.60)	0.08 (–1.62, 1.78)	.93	.68
Body mass index (kg/m^2)					
12 weeks	–0.93 (1.31)	–1.11 (1.52)	0.18 (–0.14, 0.50)	.28	
24 weeks	–0.97 (1.45)	–1.24 (2.00)	0.28 (–0.12, 0.68)	.17	.20
Resting heart rate (bpm)					
12 weeks	–1.33 (7.12)	–2.35 (6.49)	1.20 (–0.23, 2.64)	.10	
24 weeks	–1.30 (7.15)	–2.78 (6.86)	1.64 (0.12, 3.15)	.03	.03
Waist circumference at umbilicus (cm)					
12 weeks	–3.37 (4.52)	–3.57 (5.31)	0.20 (–0.93, 1.33)	.73	
24 weeks	–3.81 (5.17)	–4.93 (6.73)	1.12 (–0.25, 2.48)	.11	.26
Waist circumference at narrowest point (cm)					
12 weeks	–2.48 (4.27)	–3.37 (4.80)	0.90 (–0.13, 1.92)	.09	
24 weeks	–2.73 (4.26)	–3.79 (6.19)	1.08 (–0.13, 2.29)	.08	.06
Waist-to-height ratio at umbilicus					
12 weeks	–0.02 (0.03)	–0.02 (0.03)	0.00 (–0.01, 0.01)	.86	
24 weeks	–0.02 (0.03)	–0.03 (0.04)	0.01 (–0.00, 0.01)	.15	.35
Waist-to-height ratio at narrowest point					
12 weeks	–0.01 (0.02)	–0.02 (0.03)	0.00 (–0.00, 0.01)	.11	
24 weeks	–0.02 (0.02)	–0.02 (0.04)	0.01 (–0.00, 0.01)	.12	.09
Total serum cholesterol (mmol/L)					
12 weeks	–0.08 (0.50)	–0.19 (0.58)	0.12 (–0.00, 0.24)	.06	
24 weeks	0.01 (0.44)	–0.06 (0.58)	0.08 (–0.04, 0.19)	.20	.06

XSL•FO
RenderX

Characteristic and follow-up time	Treatment group Mean change (SD)		Absolute difference between groups LSM (95% CI)	P values for group effect	
	Basic	Enhanced	Enhanced vs basic	Difference at follow-up	Difference between groups
LDL cholesterol (mmol/L)					
12 weeks	–0.00 (0.45)	–0.07 (0.46)	0.09 (–0.02, 0.19)	.11	
24 weeks	0.02 (0.39)	–0.02 (0.46)	0.06 (–0.04, 0.16)	.27	.10
HDL cholesterol (mmol/L)					
12 weeks	–0.00 (0.14)	–0.01 (0.16)	0.00 (–0.03, 0.04)	.93	
24 weeks	0.02 (0.10)	0.02 (0.16)	0.01 (–0.02, 0.04)	.68	.86
Triglycerides (mmol/L)					
12 weeks	–0.16 (0.63)	–0.26 (0.66)	0.05 (–0.08, 0.18)	.48	
24 weeks	–0.08 (0.43)	–0.21 (0.65)	0.08 (–0.04, 0.19)	.18	.27
LDL to HDL ratio					
12 weeks	0.00 (0.40)	–0.04 (0.35)	0.06 (–0.03, 0.15)	.21	
24 weeks	–0.02 (0.34)	–0.04 (0.36)	0.03 (–0.06, 0.11)	.51	.23
Glucose (mmol/L)					
12 weeks	–0.22 (0.55)	–0.22 (0.55)	0.02 (–0.10, 0.14)	.78	
24 weeks	–0.02 (0.38)	–0.10 (0.50)	0.10 (0.00, 0.20)	.05	.22
Insulin (mIU/L)					
12 weeks	–1.76 (10.91)	–1.80 (5.47)	0.29 (–1.25, 1.82)	.71	
24 weeks	–2.37 (9.85)	–2.16 (5.92)	0.02 (–1.43, 1.47)	.98	.82
Physical functioning (SF36)					
12 weeks	2.48 (20.05)	4.71 (15.74)	1.14 (–2.34, 4.62)	.52	
24 weeks	3.48 (10.40)	3.38 (13.63)	0.81 (–1.58, 3.21)	.50	.90
Mental health (SF36)					
12 weeks	2.13 (15.06)	4.04 (13.43)	1.58 (–1.41, 4.57)	.30	
24 weeks	0.82 (11.74)	3.15 (23.39)	2.03 (–2.11, 6.17)	.34	.22
Total physical activity MET (min/week)					
12 weeks	215.29 (2448.3)	373.56 (2467.1)	96.19 (–424.3, 616.66)	.72	
24 weeks	136.47 (2303.8)	653.82 (2882.0)	474.17 (–112.4, 1060.7)	.11	.20
Average step count per day					
12 weeks	319.92 (2481.3)	1059.8 (3094.2)	675.22 (–11.32, 1361.8)	.05	
24 weeks	36.69 (1838.9)	485.19 (2766.5)	405.53 (–191.5, 1002.6)	.18	.05
Cognitive restraint scale					
12 weeks	1.32 (2.72)	1.74 (3.28)	0.48 (–0.15, 1.11)	.14	
24 weeks	1.32 (2.75)	1.71 (3.10)	0.43 (–0.22, 1.08)	.19	.12
Uncontrolled eating scale					
12 weeks	–1.71 (3.44)	–1.78 (3.41)	0.14 (–0.58, 0.86)	.70	
24 weeks	–1.55 (3.37)	–1.56 (3.48)	0.09 (–0.65, 0.82)	.82	.73
Emotional eating score					
12 weeks	–0.37 (1.38)	–0.48 (1.56)	0.10 (–0.22, 0.41)	.54	
24 weeks	–0.40 (1.59)	–0.57 (1.68)	0.17 (–0.19, 0.52)	.36	.38

Characteristic and follow-up time	Treatment group Mean change (SD)		Absolute difference between groups LSM (95% CI)	P values for group effect	
	Basic	Enhanced	Enhanced vs basic	Difference at follow-up	Difference between groups
Total energy intake					
12 weeks	–985.5 (2386.7)	–1248 (2458.1)	267.74 (–186.4, 721.90)	.25	
24 weeks	–787.1 (2095.4)	–761.4 (2046.2)	22.36 (–407.4, 452.10)	.92	.54

Table 5. Mean change in a range of variables from baseline to 12 weeks and baseline to 24 weeks within each treatment group and the least squares mean (LSM) difference in change between treatment groups (completers population).

Characteristic and follow-up time	Treatment group Mean change (SD)		Absolute difference between groups LSM (95% CI)	*P* values for group effect	
	Basic	Enhanced	Enhanced vs basic	Difference at follow-up	Difference between groups
Weight (kg)					
12 weeks	–3.7 (4.0)	–4.3 (4.6)	0.6 (–0.6, 1.8)	.34	
24 weeks	–4.1 (4.7)	–4.8 (6.6)	0.7 (–0.8, 2.3)	.35	.35
Percentage weight loss (%)					
12 weeks	–3.89 (4.10)	–4.59 (4.81)	0.69 (–0.54, 1.92)	.27	
24 weeks	–4.48 (4.91)	–5.18 (6.70)	0.71 (–0.88, 2.31)	.38	.27
Attained 5% weight loss (%)					
12 weeks	33.7 (47.5)	41.3 (49.4)	7.65 (–5.65, 20.94)	.26	
24 weeks	41.2 (49.5)	45.7 (50.0)	4.76 (–8.40, 17.92)	.48	.31
Systolic blood pressure (mmHg)					
12 weeks	–4.91 (12.38)	–4.91 (10.63)	0.37 (–2.48, 3.21)	.80	
24 weeks	–3.98 (12.11)	–2.95 (12.65)	0.78 (–2.15, 3.71)	.60	.65
Diastolic blood pressure (mmHg)					
12 weeks	–2.87 (10.33)	–2.86 (8.37)	0.43 (–1.72, 2.59)	.69	
24 weeks	–1.43 (10.89)	–1.25 (8.90)	0.22 (–2.14, 2.58)	.85	.91
Body mass index (kg/m^2)					
12 weeks	–1.27 (1.33)	–1.42 (1.57)	0.15 (–0.26, 0.55)	.48	
24 weeks	–1.53 (1.57)	–1.70 (2.17)	0.16 (–0.37, 0.70)	.55	.48
Resting heart rate (bpm)					
12 weeks	–1.90 (8.31)	–3.07 (7.18)	1.37 (–0.48, 3.21)	.15	
24 weeks	–2.09 (8.97)	–3.81 (7.78)	1.76 (–0.34, 3.87)	.10	.05
Waist circumference at umbilicus (cm)					
12 weeks	–4.44 (4.56)	–4.51 (5.50)	0.02 (–1.38, 1.42)	.98	
24 weeks	–6.13 (5.36)	–6.83 (7.06)	0.56 (–1.16, 2.29)	.52	.64
Waist circumference at narrowest point (cm)					
12 weeks	–3.18 (4.54)	–4.28 (5.01)	1.05 (–0.24, 2.34)	.11	
24 weeks	–4.39 (4.68)	–5.25 (6.75)	0.82 (–0.81, 2.44)	.32	.15
Waist-to-height ratio at umbilicus					
12 weeks	–0.03 (0.03)	–0.03 (0.03)	0.00 (–0.01, 0.01)	.91	
24 weeks	–0.04 (0.03)	–0.04 (0.04)	0.00 (–0.01, 0.01)	.65	.75
Waist-to-height ratio at narrowest point					
12 weeks	–0.02 (0.03)	–0.02 (0.03)	0.01 (–0.00, 0.01)	.14	
24 weeks	–0.03 (0.03)	–0.03 (0.04)	0.00 (–0.01, 0.01)	.42	.19
Total serum cholesterol (mmol/L)					
12 weeks	–0.12 (0.54)	–0.23 (0.65)	0.12 (–0.04, 0.28)	.14	
24 weeks	0.02 (0.57)	–0.09 (0.70)	0.12 (–0.06, 0.30)	.20	.09

XSL•FO
RenderX

Characteristic and follow-up time	Treatment group Mean change (SD)		Absolute difference between groups LSM (95% CI)	P values for group effect	
	Basic	Enhanced	Enhanced vs basic	Difference at follow-up	Difference between groups
LDL cholesterol (mmol/L)					
12 weeks	–0.02 (0.45)	–0.08 (0.53)	0.08 (–0.06, 0.22)	.27	
24 weeks	0.04 (0.52)	–0.02 (0.57)	0.09 (–0.07, 0.26)	.27	.26
HDL cholesterol (mmol/L)					
12 weeks	–0.01 (0.14)	–0.01 (0.18)	0.01 (–0.04, 0.05)	.76	
24 weeks	0.03 (0.13)	0.04 (0.19)	0.00 (–0.04, 0.05)	.86	.71
Triglycerides (mmol/L)					
12 weeks	–0.21 (0.64)	–0.34 (0.73)	0.06 (–0.10, 0.22)	.47	
24 weeks	–0.14 (0.56)	–0.30 (0.77)	0.07 (–0.09, 0.24)	.38	.21
LDL to HDL ratio					
12 weeks	–0.02 (0.44)	–0.05 (0.40)	0.05 (–0.07, 0.18)	.38	
24 weeks	–0.04 (0.46)	–0.07 (0.44)	0.04 (–0.10, 0.17)	.61	.43
Glucose (mmol/L)					
12 weeks	–0.31 (0.64)	–0.27 (0.61)	0.00 (–0.16, 0.16)	.99	
24 weeks	–0.03 (0.50)	–0.15 (0.61)	0.15 (–0.00, 0.29)	.05	.35
Insulin (mIU/L)					
12 weeks	–2.69 (13.26)	–2.25 (6.14)	0.48 (–1.13, 2.08)	.56	
24 weeks	–4.03 (12.62)	–3.15 (6.95)	0.20 (–1.34, 1.73)	.80	.61
Physical functioning (SF36)					
12 weeks	3.54 (24.91)	5.21 (16.40)	0.57 (–3.93, 5.08)	.80	
24 weeks	5.57 (12.74)	4.78 (15.93)	1.25 (–1.89, 4.40)	.43	.88
Mental health (SF36)					
12 weeks	2.36 (18.09)	4.88 (14.97)	1.39 (–2.51, 5.28)	.48	
24 weeks	1.31 (14.87)	4.38 (27.63)	1.87 (–4.18, 7.93)	.54	.53
Total physical activity MET (min/week)					
12 weeks	164.57 (2799.8)	432.99 (2735.8)	214.87 (–403.0, 832.76)	.49	
24 weeks	230.41 (2997.8)	932.18 (3408.2)	491.04 (–381.0, 1363.0)	.27	.28
Average step count per day					
12 weeks	418.22 (3102.2)	1598.7 (3600.4)	1139.3 (118.76, 2159.9)	.03	
24 weeks	91.73 (2926.8)	808.65 (3543.6)	587.05 (–604.7, 1778.8)	.33	.06
Cognitive restraint scale					
12 weeks	2.11 (2.94)	2.29 (3.48)	0.01 (–0.78, 0.79)	.98	
24 weeks	2.13 (3.24)	2.44 (3.46)	0.19 (–0.69, 1.07)	.66	.66
Uncontrolled eating scale					
12 weeks	–2.60 (3.82)	–2.26 (3.76)	0.12 (–0.81, 1.05)	.81	
24 weeks	–2.46 (3.98)	–2.20 (3.96)	0.07 (–0.92, 1.07)	.89	.96
Emotional eating score					
12 weeks	–0.53 (1.61)	–0.59 (1.75)	0.06 (–0.37, 0.48)	.79	
24 weeks	–0.63 (1.98)	–0.79 (1.94)	0.14 (–0.36, 0.64)	.58	.58

Characteristic and follow-up time	Treatment group Mean change (SD)		Absolute difference between groups LSM (95% CI)	P values for group effect	
	Basic	Enhanced	Enhanced vs basic	Difference at follow-up	Difference between groups
Total energy intake					
12 weeks	–1480 (2695.4)	–1467 (2428.2)	90.91 (–462.8, 644.63)	.75	
24 weeks	–1285 (2559.6)	–1047 (2338.7)	208.94 (–379.6, 797.48)	.48	.82

Table 6. Mean change in total website usage for the completers population and the ITT with LOCF from baseline to 12 weeks and baseline to 24 weeks within each treatment group.

Characteristic and follow-up time	Treatment group Mean change (SD)		Absolute difference between groups LSM (95% CI)	P values for group effect	
	Basic	Enhanced	Enhanced vs basic	Difference at follow-up	Difference between groups
Intention-to-treat					
Total website was usage since baseline (days)					
12 weeks	24.6 (25.5)	34.1 (28.1)	9.45 (3.34, 15.56)	.002	
24 weeks	31.8 (33.9)	43.1 (34.0)	12.47 (4.73, 20.20)	.002	.002
Completers					
Total website was usage since baseline (days)					
12 weeks	29.7 (26.6)	38.7 (28.5)	9.2 (–1.9, 16.5)	.01	
24 weeks	42.0 (36.1)	49.8 (33.3)	8.1 (–3.3, 19.4)	.16	.02

Discussion

Principal Findings

The aim of the study was to evaluate whether overweight and obese adults randomized to a commercial Web-based weight-loss program providing greater social support and more personalized feedback achieved a greater reduction in BMI and increased usage of program features compared to those randomized to a standard version of the program. We found no differences in weight loss or most of the secondary health outcomes between the basic and enhanced features versions of the Web-based weight-loss program after 24 weeks, despite previous reports that provision of enhanced features within Web-based formats does enhance weight-loss outcomes [5]. Mean weight loss in the current study ranged from 2 to 3 kg after 12 weeks and 3 to 4 kg after 24 weeks across both intervention groups. Both the magnitude of weight loss, and the continuance of weight loss from 12 to 24 weeks highlights that both versions are promising at the population level. The results also compare favorably to previous Web-based studies. Only 4 of 7 studies were deemed effective with a mean weight loss ≥5% [24-28]. When we examined those who achieved ≥5% weight loss, success was strongly associated with website usage, indicating that strategies to improve website usage may be beneficial to weight loss outcomes. In this regard, some aspects of the enhanced program features may be valuable because the enhanced group had a significantly lower dropout rate and greater participant engagement.

In the current study, the basic and enhanced versions may have produced similar weight loss because several of these components were similar (self-monitoring, social support, structured program) or absent (eg, counselor feedback) in both versions. Khaylis et al [29] reviewed technology-based weight-loss intervention studies and identified 5 factors that may contribute to successful weight loss: use of a structured program, self-monitoring, social support, use of an individually tailored program, and counselor feedback and communication. Although semipersonalized system-generated feedback and an escalating reminder scale to begin was provided in the enhanced group, the report may not have been specific enough to help them further improve their dietary intake, physical activity, and log-ins. The contact may have been viewed as too much contact and, therefore, contributed to nonusage. In the current study, the basic version of the Web-based weight-loss program proved effective, supported by the 0.9 kg/m^2 reduction in BMI (2.7 kg) at 12 weeks and the 1.1 kg/m^2 BMI reduction (3.3 kg) after 24 weeks. This degree of weight loss is similar to that in the enhanced arm of older Web-based trials. For example, in 2001, Tate et al [28] reported 3- and 6-month weight losses of -3.2 kg and -2.9 kg, respectively, in 46 adults in the enhanced arm of an Internet weight-loss trial compared to -1.0 kg and -1.3 kg for the basic group, whereas in 2006, Rothert et al [8] reported that for 1475 adults participating in the tailored (enhanced) feedback arm, the mean weight losses at 3 and 6 months were 0.8% and 0.9% body weight compared to -0.4% in the basic information-only Internet program at both 3 and 6 months. A

recent 2010 study by Wing et al [6] did report significantly greater weight loss in an enhanced Web program (-3.1 kg) compared to a basic version (-1.2 kg), with mean weight loss for the enhanced similar to the basic program in the current study. Another study by Webber et al [10] reported greater weight loss using Web-based programs than the current study. Although the weight loss between a basic (minimal contact) and enhanced version of an Internet program was not significantly different, both groups achieved substantial weight losses of 5.2 kg (minimal) and 3.7 kg (enhanced) after 16 weeks [10]. However, it is difficult to compare and discern how different Web-based features may influence outcomes. Having standardized ways of describing or reporting enhanced program features would assist in making comparisons across studies and, over time, could help with identifying the set of program components that may optimize weight loss using Web-based programs. Future programs may need to segment the target population to improve feedback tailoring to specific user groups as a strategy to avoid website discontinuity, particularly in relation to some age, sex, or BMI subgroups [30]. Some groups may not need this more extensive feedback and it would be useful to identify who they are. Although there were some minor differences in outcomes across categories of age, BMI, and sex, a Web-based program may potentially benefit specific groups of program users [31,32].

In the completers population of the current study, significant correlations were found between total website usage and the percentage weight loss at 12 and 24 weeks. Participants who achieved clinically important weight loss (≥5%) at either time point used the website almost 4 times more than those who were not successful (<5% weight loss). Further, the website log-ins among those deemed successful was substantially greater than the number of log-ins reported for the enhanced group. Those with successful weight loss logged in 2 to 3 times per week, compared to just once or twice a week for those randomized to the enhanced group and less than once a week for those randomized to the basic group.

The correlation between number of log-ins and weight loss was moderate across all study participants and there was also no between-group difference. This suggests that although being allocated to the enhanced program did facilitate more frequent website log-ins, provision of additional features is not enough to facilitate greater engagement and weight loss. Further research examining which combination of website features optimize

program use and reduce attrition are needed. Based on the current study, future modifications to the enhanced program would need to achieve a 50% increase in the number of participant log-ins than that in the current study. This would mean getting participants to use the program at least 2 to 3 times per week as a strategy to facilitate clinically important ≥5% weight loss. Establishing and testing these targets could ease the burden and fatigue associated with program usage targets that are not achievable or sustainable. Although between-groups differences might typically be explained by confounders such as energy intake and physical activity, these were not different between groups. It is more likely that differential use of social support features, including blogs, forums, and chat rooms, explain the between-group difference and this requires further research. We cannot tell whether the reminders schedule to log in and use of program features in the current study was the key driver of this and this also needs to be examined in future studies.

Limitations

A limitation of the current study is that it did not have a waitlist control group at 6 months. However, this was not required to answer the research question. Attrition reduced the power to detect significant differences between groups, particularly for the secondary outcomes; however, there were a few trends suggesting that better retention would not have changed the outcomes in any substantial way. The strengths of this study include the use of an RCT, large sample size, use of blinded assessors, and the comparison of the effectiveness of the 2 versions of the weight-loss programs up to 24 weeks. Further, few commercial Web-based programs have been subjected to evaluation by RCT, with none previously conducted in Australia. Importantly, this study has demonstrated the efficacy of a commercial Web-based weight-loss program in achieving clinically important weight loss.

Conclusions

In conclusion, commercial Web-based weight-loss programs can be effective at achieving clinically meaningful weight loss up to 24 weeks. Although adding enhanced features that provide additional feedback, reminders, and social support promotes greater retention and engagement, it does not necessarily increase weight loss substantially. Further research into Web-based features that optimize website usage, program engagement, and weight-loss success is warranted.

XSL•FO

RenderX

Conflicts of Interest

CEC has been a nutrition consultant to SP Health Co M Hutchesson (nee Neve) was funded by a Penn Health postdoctoral fellowship. All other authors declare that they have no competing interests.

Multimedia Appendix 1

CONSORT-eHEALTH checklist V1.6.2 [15].

[PDF File (Adobe PDF File), 1MB - jmir_v15i7e140_app1.pdf]

Multimedia Appendix 2

Sample weekly report_5.

[JPG File, 772KB - jmir_v15i7e140_app2.jpg]

Multimedia Appendix 3

Sample weekly report_5a.

[JPG File, 787KB - jmir_v15i7e140_app3.jpg]

References

1. Ogden CL, Yanovski SZ, Carroll MD, Flegal KM. The epidemiology of obesity. Gastroenterology 2007 May;132(6):2087-2102. [doi: 10.1053/j.gastro.2007.03.052] [Medline: 17498505]
2. Smith A. Home broadband. Washington, DC: Pew Internet & American Life Project; 2010 Aug 11. URL: http://pewinternet.org/Reports/2010/Home-Broadband-2010.aspx [accessed 2013-06-30] [WebCite Cache ID 6HmGW8lmF]
3. Australian Bureau of Statistics. Canberra, Australia: ABS; 2010. 8153.0-Internet Activity, Australia, Jun 2010 URL: http://www.abs.gov.au/AUSSTATS/abs@.nsf/Lookup/8153.0Main+Features1Jun%202010?OpenDocument [accessed 2013-06-30] [WebCite Cache ID 6HmGeFS05]
4. Fox S, Jones S. The social life of health information: Americans' pursuit of health takes place within a widening network of both online and offline sources. Washington, DC: Pew Internet & American Life Project; 2009 Jun. URL: http://www.pewinternet.org/~/media//Files/Reports/2009/PIP_Health_2009.pdf [accessed 2013-06-30] [WebCite Cache ID 6HmGsMeVS]
5. Neve M, Morgan PJ, Jones PR, Collins CE. Effectiveness of web-based interventions in achieving weight loss and weight loss maintenance in overweight and obese adults: a systematic review with meta-analysis. Obes Rev 2010 Apr;11(4):306-321. [doi: 10.1111/j.1467-789X.2009.00646.x] [Medline: 19754633]
6. Wing RR, Crane MM, Thomas JG, Kumar R, Weinberg B. Improving weight loss outcomes of community interventions by incorporating behavioral strategies. Am J Public Health 2010 Dec;100(12):2513-2519. [doi: 10.2105/AJPH.2009.183616] [Medline: 20966375]
7. Pullen CH, Hageman PA, Boeckner L, Walker SN, Oberdorfer MK. Feasibility of Internet-delivered weight loss interventions among rural women ages 50-69. J Geriatr Phys Ther 2008;31(3):105-112. [Medline: 19856616]
8. Rothert K, Strecher VJ, Doyle LA, Caplan WM, Joyce JS, Jimison HB, et al. Web-based weight management programs in an integrated health care setting: a randomized, controlled trial. Obesity (Silver Spring) 2006 Feb;14(2):266-272. [doi: 10.1038/oby.2006.34] [Medline: 16571852]
9. Krukowski RA, Harvey-Berino J, Ashikaga T, Thomas CS, Micco N. Internet-based weight control: the relationship between web features and weight loss. Telemed J E Health 2008 Oct;14(8):775-782 [FREE Full text] [doi: 10.1089/tmj.2007.0132] [Medline: 18954247]
10. Webber KH, Tate DF, Michael Bowling J. A randomized comparison of two motivationally enhanced Internet behavioral weight loss programs. Behav Res Ther 2008 Sep;46(9):1090-1095. [doi: 10.1016/j.brat.2008.06.008] [Medline: 18675402]
11. Tsai AG, Wadden TA. Systematic review: an evaluation of major commercial weight loss programs in the United States. Ann Intern Med 2005 Jan 4;142(1):56-66. [Medline: 15630109]
12. Collins CE, Morgan PJ, Jones P, Fletcher K, Martin J, Aguiar EJ, et al. Evaluation of a commercial web-based weight loss and weight loss maintenance program in overweight and obese adults: a randomized controlled trial. BMC Public Health 2010;10:669 [FREE Full text] [doi: 10.1186/1471-2458-10-669] [Medline: 21047432]
13. Collins CE, Morgan PJ, Jones P, Fletcher K, Martin J, Aguiar EJ, et al. A 12-week commercial web-based weight-loss program for overweight and obese adults: randomized controlled trial comparing basic versus enhanced features. J Med Internet Res 2012;14(2):e57 [FREE Full text] [doi: 10.2196/jmir.1980] [Medline: 22555246]
14. Norton K. Sports Medicine Australia. Canberra, Australia: Australian Department of Health and Ageing; 2005. Sports Medicine Australia pre-exercise screening system URL: http://sma.org.au/wp-content/uploads/2009/05/new_pre_screening.pdf [accessed 2013-06-30] [WebCite Cache ID 6HmH79t2Q]

15. Eysenbach G, CONSORT-EHEALTH Group. 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]

16. National Health and Medical Research Council. Nutrient Reference Values for Australia and New Zealand including Recommended Dietary Intakes. Canberra, Australia: Department of Health and Ageing; 2006.

17. Bandura A. Social Foundations of Thought and Action: A Social Cognitive Theory. Englewood Cliffs, NJ: Prentice-Hall; 1986.

18. Watson J, Collins C, Guest M, Pezdirc K, Burrows T, Huxley S. PP240: Evaluation of an adult food frequency questionnaire and its associated diet quality score. In: The University of Newcastle.: FAO; 2012 Presented at: International Conference on Diet Activity and Methods; May 14-17, 2012; Rome, Italy p. 198.

19. Australia New Zealand Food Authority. AUSNUT: Australian Food and Nutrient database 1999 for estimation of dietary intake. Canberra, Australia: Australian Government Publishing Service; 1999.

20. Karlsson J, Persson LO, Sjöström L, Sullivan M. Psychometric properties and factor structure of the Three-Factor Eating Questionnaire (TFEQ) in obese men and women. Results from the Swedish Obese Subjects (SOS) study. Int J Obes Relat Metab Disord 2000 Dec;24(12):1715-1725. [Medline: 11126230]

21. Ware J, Kosinski M, Dewey J. How to Score Version 2 of the SF-36 Health Survey. Lincoln, RI: Quality Metric Incorporated; 2001.

22. Guidelines for data processing and analysis of the International Physical Activity Questionnaire (IPAQ)- Short and long forms.: IPAQ; 2005. URL: http://www.ipaq.ki.se/scoring.pdf [accessed 2013-07-01] [WebCite Cache ID 6HmK3JqUA]

23. Christian JG, Tsai AG, Bessesen DH. Interpreting weight losses from lifestyle modification trials: using categorical data. Int J Obes (Lond) 2010 Jan;34(1):207-209 [FREE Full text] [doi: 10.1038/ijo.2009.213] [Medline: 19823184]

24. Gold BC, Burke S, Pintauro S, Buzzell P, Harvey-Berino J. Weight loss on the web: A pilot study comparing a structured behavioral intervention to a commercial program. Obesity (Silver Spring) 2007 Jan;15(1):155-164. [doi: 10.1038/oby.2007.520] [Medline: 17228043]

25. Harvey-Berino J, Pintauro S, Buzzell P, Gold EC. Effect of internet support on the long-term maintenance of weight loss. Obes Res 2004 Feb;12(2):320-329. [doi: 10.1038/oby.2004.40] [Medline: 14981225]

26. Polzien KM, Jakicic JM, Tate DF, Otto AD. The efficacy of a technology-based system in a short-term behavioral weight loss intervention. Obesity (Silver Spring) 2007 Apr;15(4):825-830. [doi: 10.1038/oby.2007.584] [Medline: 17426316]

27. Tate DF, Jackvony EH, Wing RR. A randomized trial comparing human e-mail counseling, computer-automated tailored counseling, and no counseling in an Internet weight loss program. Arch Intern Med 2006;166(15):1620-1625. [doi: 10.1001/archinte.166.15.1620] [Medline: 16908795]

28. Tate DF, Wing RR, Winett RA. Using Internet technology to deliver a behavioral weight loss program. JAMA 2001 Mar 7;285(9):1172-1177. [Medline: 11231746]

29. Khaylis A, Yiaslas T, Bergstrom J, Gore-Felton C. A review of efficacious technology-based weight-loss interventions: five key components. Telemed J E Health 2010 Nov;16(9):931-938 [FREE Full text] [doi: 10.1089/tmj.2010.0065] [Medline: 21091286]

30. Lehto T, Oinas-Kukkonen H. Persuasive Features in Six Weight Loss Websites: A Qualitative Evaluation. Berlin/Heidelberg: Springer; 2010 Presented at: Persuasive Technology, 5th International Conference, PERSUASIVE; June 7-10, 2010; Copenhagen, Denmark p. 162-173.

31. Neve MJ, Morgan PJ, Collins CE. Behavioural factors related with successful weight loss 15 months post-enrolment in a commercial web-based weight-loss programme. Public Health Nutr 2012 Jul;15(7):1299-1309. [doi: 10.1017/S1368980011003090] [Medline: 22122973]

32. Neve M, Morgan PJ, Collins CE. Weight change in a commercial web-based weight loss program and its association with website use: cohort study. J Med Internet Res 2011;13(4):e83 [FREE Full text] [doi: 10.2196/jmir.1756] [Medline: 21993231]

Abbreviations

AES: Australian Eating Survey
ANOVA: analysis of variance
ANZCTR: Australian New Zealand Clinical Trials Registry
BMI: body mass index
BOCF: baseline observation carried forward
CONSORT: Consolidated Standards of Reporting Trials
FFQ: food-frequency questionnaire
GLMM: generalized linear mixed model
HDL: high-density lipoprotein
IPAQ-SF: International Physical Activity Questionnaire-short form
ITT: intention-to-treat
LDL: low-density lipoprotein

XSL•FO
RenderX

LOCF: last observation carried forward
LSM: least squares mean
MET: total metabolic equivalent
RCT: randomized controlled trial
SMS: short message service
TFEQ-R18: Three-Factor Eating Questionnaire-R18

Original Paper

The Effectiveness of a Web-Based Personalized Feedback and Social Norms Alcohol Intervention on United Kingdom University Students: Randomized Controlled Trial

Bridgette M Bewick[1], BA, MA (Hons), PhD; Robert M West[2]; Michael Barkham[3]; Brendan Mulhern[4]; Robert Marlow[5]; Gemma Traviss[1]; Andrew J Hill[1]

[1]Academic Unit of Psychiatry and Behavioural Sciences, Leeds Institute of Health Sciences, School of Medicine, University of Leeds, Leeds, United Kingdom

[2]Division of Biostatistics, University of Leeds, Leeds, United Kingdom

[3]Centre for Psychological Services Research, University of Sheffield, Sheffield, United Kingdom

[4]School of Health and Related Research, University of Sheffield, Sheffield, United Kingdom

[5]School of Computing, University of Leeds, Leeds, United Kingdom

Corresponding Author:
Bridgette M Bewick, BA, MA (Hons), PhD
Academic Unit of Psychiatry and Behavioural Sciences, Leeds Institute of Health Sciences
School of Medicine
University of Leeds
Charles Thackrah Building
101 Clarendon Road
Leeds, LS2 9LJ
United Kingdom
Phone: 44 1133430809
Email: B.M.Bewick@leeds.ac.uk

Abstract

Background: Alcohol consumption in the student population continues to be cause for concern. Building on the established evidence base for traditional brief interventions, interventions using the Internet as a mode of delivery are being developed. Published evidence of replication of initial findings and ongoing development and modification of Web-based personalized feedback interventions for student alcohol use is relatively rare. The current paper reports on the replication of the initial Unitcheck feasibility trial.

Objective: To evaluate the effectiveness of Unitcheck, a Web-based intervention that provides instant personalized feedback on alcohol consumption. It was hypothesized that use of Unitcheck would be associated with a reduction in alcohol consumption.

Methods: A randomized control trial with two arms (control=assessment only; intervention=fully automated personalized feedback delivered using a Web-based intervention). The intervention was available week 1 through to week 15. Students at a UK university who were completing a university-wide annual student union electronic survey were invited to participate in the current study. Participants (n=1618) were stratified by sex, age group, year of study, self-reported alcohol consumption, then randomly assigned to one of the two arms, and invited to participate in the current trial. Participants were not blind to allocation. In total, n=1478 (n=723 intervention, n=755 control) participants accepted the invitation. Of these, 70% were female, the age ranged from 17-50 years old, and 88% were white/white British. Data were collected electronically via two websites: one for each treatment arm. Participants completed assessments at weeks 1, 16, and 34. Assessment included CAGE, a 7-day retrospective drinking diary, and drinks consumed per drinking occasion.

Results: The regression model predicted a monitoring effect, with participants who completed assessments reducing alcohol consumption over the final week. Further reductions were predicted for those allocated to receive the intervention, and additional reductions were predicted as the number of visits to the intervention website increased.

Conclusions: Unitcheck can reduce the amount of alcohol consumed, and the reduction can be sustained in the medium term (ie, 19 weeks after intervention was withdrawn). The findings suggest self-monitoring is an active ingredient to Web-based personalized feedback.

XSL•FO
RenderX

Introduction

Alcohol consumption in the student population continues to be cause for concern [1-3]. Heavy episodic or binge drinking is prevalent in this population (eg, [4]), increasing the risk of engaging in risky, illegal, and violent behaviors [5-7]. In addition to the immediate personal and societal costs associated with alcohol misuse, heavy consumption during college and university is predictive of alcohol dependence in later life. Despite this, help-seeking behavior for alcohol use is low in the student population [8], meaning relatively few students access the traditional support services available.

Building on the established evidence base for traditional brief interventions, interventions using the Internet as a mode of delivery are being developed. Such developments have potential to aid early identification and reach their targets on a population level. Emerging evidence suggests that interventions targeted at eHealth care systems aimed at reducing harmful alcohol use that are implemented as part of a wider health care system can be cost-effective [9]. There is evidence that Internet interventions with and without therapist support can provide cost-effective behavior change with those drinking at harmful levels [10]. The potential for eHealth interventions to intervene early and engage non-help-seeking individuals means eHealth solutions for providing personalized feedback to the general population hold the potential to increase effectiveness and cost-effectiveness of public health interventions. The cost-effectiveness of this approach requires further investigation. But the ability to engage individuals in personalized feedback on a population basis combined with an ability to enable confidential access at a time convenient to the user makes electronic delivery of interventions attractive.

There is evidence that Web-based interventions that provide personalized feedback and incorporate social norms information can be effective in moderating alcohol use [11-14]. Conventional approaches to alcohol and drug health education were based upon an assumed lack of knowledge concerning the risks associated with drinking alcohol. These risk-focused campaigns are increasingly viewed as ineffectual [15]. In particular, it is acknowledged that risk-based campaigns may be dismissed by the target population due to the relatively low occurrence of risk events within the general population [16].

The social norms approach recognizes that people tend to overestimate the alcohol consumption of others and that these misperceptions predict heavier alcohol use [17,18]. There is growing evidence that interventions that include instant personalized social norms feedback can reduce alcohol consumption [19]. Recent reviews, however, have pointed to inconsistencies in reported effectiveness and efficacy. These differences can be explained by weaknesses in the methodological quality of some evaluations [19-22] and by differences in the immediacy of feedback [23]. Reviews have highlighted the need for further studies that utilize rigorous research designs [20-22] and that include longer follow-up data [21,24].

Published evidence of replication of initial findings and ongoing development and modification of Web-based personalized feedback interventions for student alcohol use is relatively rare. Exceptions include the body of work investigating e-CHUG [25,26], Unitcheck [12,27], and developments following the e-SBI pilot trial conducted by Kypri [11,28,29].

The current paper reports on the replication of the initial Unitcheck feasibility trial [27]. The feasibility randomized controlled trial (RCT) recruited 506 participants from a single UK university. After completing an online assessment, intervention participants received brief electronic personalized feedback. The intervention was available over a 12-week period, and participants could log on at any time and receive instant feedback. The trial reported a significant difference in Time 1 (week 1) to Time 2 (week 12) alcohol consumed per occasion. However, no significant difference was found for units of alcohol consumed over the previous week (1 UK unit=10 mL ethanol). As a feasibility study, the trial had a number of methodological shortcomings. No information was collected on daily alcohol intake so it was not possible to examine possible intervention effects on drinks per day over the previous week. As data were collected at only two time points (week 1 and week 12), the trial could say nothing about the short- to long-term effect of the intervention. There is a need for additional research that seeks to replicate, and understand further, initial findings and how intervention developments affect outcome. The current study sought to address these limitations and to evaluate the intervention in a larger sample.

Accordingly, the aim of the current study was to evaluate the effectiveness of Unitcheck, and the hypothesis tested was that use of Unitcheck would be associated with a reduction in alcohol consumption.

Methods

Setting

The study was an RCT conducted at the University of Leeds, a UK university located in the Yorkshire and Humber region of England. During the time that this study was undertaken, not all non-clinical RCTs were expected to be registered (Multimedia Appendix 2).

Procedure and Participants

In January 2007, students completing a university-wide annual student union electronic survey (n=4528) were invited to participate in a study investigating student alcohol consumption. Students who registered their interest, gave initial online consent, and provided data at baseline indicating they were a consumer of alcohol (n=1618; Time 0=T0) were invited to participate in the current study (see Figure 1). Participants were

asked to complete online assessments at week 1 (Time 1=T1), week 16 (Time 2=T2), and week 34 (Time 3=T3). Those allocated to receive the intervention had access to the website from week 1 to week 15. Control participants completed all self-assessments using an online survey (created using Bristol Online Survey), and intervention participants completed T1 and T2 assessments via the Unitcheck intervention website. T3 self-assessments were completed using an online survey (created using Bristol Online Survey). Participation was anonymous. Response rates at each time point were as follows: Time 1, 65% (n=1049); Time 2, 46% (n=743); and Time 3, 40% (n=644). The intervention was accessed by 74% (n=535) of participants allocated to the intervention condition.

As an incentive to participate in the study, participants received university printer credits depending on their level of participation, with the maximum total amount (150 printer credits valued at £1.50) being given to individuals in the intervention condition who completed T1 (week 1), T2, and T3 assessments and also visited the site during week 7. The maximum total amount available to control participants was valued at £1.25.

The study was approved by Leeds East NHS Research Ethics Committee.

Figure 1. Participant flow through the trial.
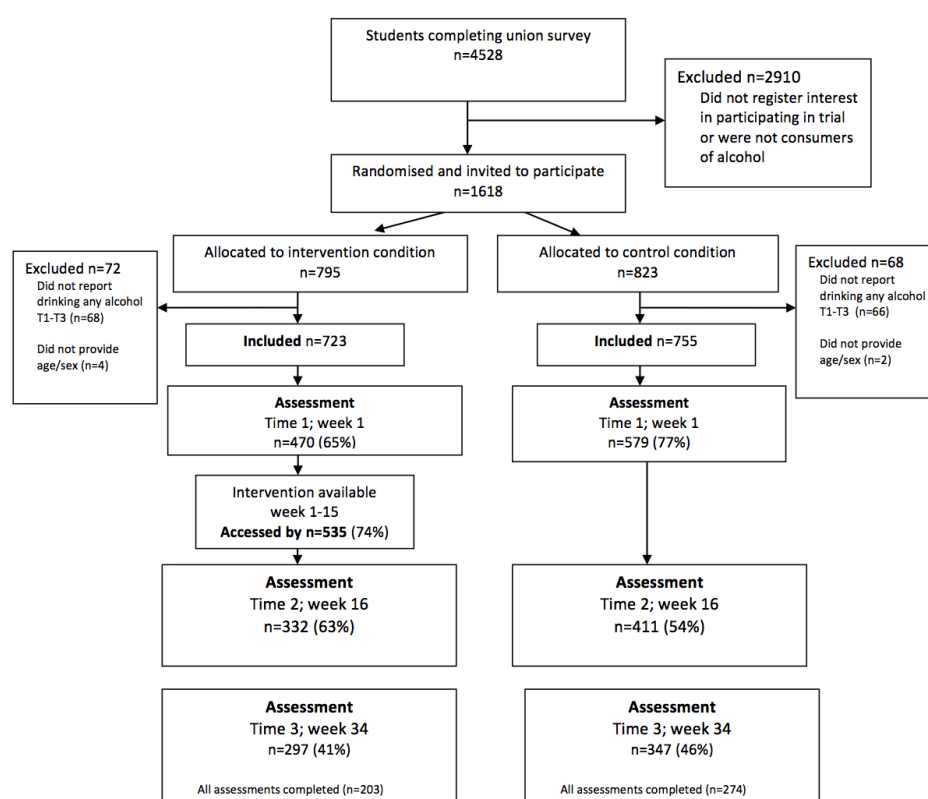
Research Design

The study was an RCT with two arms: a control arm (assessment only) and an intervention arm (access to a website providing instant personalized feedback). Participants were stratified by sex, age group, year of study, self-reported weekly alcohol consumption (classified by department of health risk level) and randomly assigned (by a researcher not involved in the current study) to one of the two arms. Participants were not blind to allocation.

Data were collected electronically via two websites: one for each treatment arm. Both websites included the same questions

presented in the same order. Contact with participants was by email, and at each stage participants received a standardized message inviting them to participate in the study. Each message included a direct link to the appropriate Web-based survey. Those who did not initially respond to the study were sent an email reminder once a week for up to 3 weeks. All participants were informed that they would be randomly allocated to a control (ie, assessment only) or an intervention arm. Immediately after completing the T1 assessment intervention participants received personalized feedback and social norms information. Intervention participants had access to the intervention website between T1 and T2 (15 weeks), and there were no restrictions placed on the number of visits they could

make to the site. Those in the intervention arm received an additional email invitation to visit the intervention website at week 7.

Sample Size

The distribution of alcohol units consumed over the last week is skewed; transformed data is closer to being normally distributed. This adds distributional validity to our modeling. From previous work we ascertained that the average natural logarithm of the number of units of alcohol consumed over the last week plus 1 for students is approximately 1.3 with a standard deviation of 0.58 and, hence, a variance of 0.34. Sample size determination is based on a matched-pairs t test. A change in natural logarithm plus 1 over the intervention period will therefore have a variance of less than 0.68 (2 times 0.345, or the variance of first measure plus the variance of the second). We have taken it to be equal to 0.49 (ie, 0.7^2).

The difference in the change between two treatment arms might be tested with a t test where the relevant standard deviation is 0.7. A suitable difference in change in the natural logarithm of the number of units consumed over the last week plus 1 was taken as 0.2, so that we sought a standardized difference of 0.29. For a significance level of alpha equal to 0.05 and 90% power, a sample size of 258 participants per treatment arm was required. To allow for attrition, we aimed to recruit at least 688 participants in total.

A change of 0.2 in log(units+1) corresponds to a change in units of around 4-5 units at the average level of drinking of 21 units per week.

Assessments

The CAGE is an assessment that was widely used as a screening tool for alcohol use disorders [30,31]. It consists of four items: (1) have you ever thought about Cutting down on your drinking, (2) do you ever get Annoyed at criticism of your drinking, (3) do you ever feel Guilty about your drinking, and (4) do you ever have a drink in the morning (an Eye-opener). Scoring positively on two or more of the items indicates problem drinking. The CAGE has previously been used within college populations [32] and has good internal consistency (alpha values between 0.52 and 0.90; [33]).

Participants were asked to report the typical number of alcoholic drinks they usually consume per drinking occasion (collected T1-T3) and how many alcoholic drinks they consumed over the last week (collected T0-T3) using a 7-day retrospective drinking diary. This method is recommended for use within samples that consume alcohol regularly [34]. The diary included a list of common alcoholic beverages and for each day of the last week/per average occasion asked participants to indicate how many of each drink they had consumed over the relevant time period. The number of alcoholic drinks consumed was then converted into UK units of alcohol consumed (1 unit=10 mL ethanol). As a result of completing the drinking diary, the number of days of alcohol consumption per week was also recorded. Weekly unit consumption was subsequently categorized according to UK government guidelines [35], namely, within recommended weekly guidelines (female 0-14 units, male 0-21 units), hazardous weekly consumption (female

15-35 units, male 22-50 units), and harmful weekly consumption (female >35 units, male >50 units). For the purposes of providing feedback, those drinking at hazardous levels were further split into two categories (female 14-21 units, male 22-28 units and female 22-35 units, male 29-50 units).

In order to assess risk behavior, participants were asked if, in the last 12 months they had experienced the following: injury to self accidentally, deliberate self-harm, injury caused by others who have been drinking, damage to property while drinking, and sexual intercourse when they ordinarily would not.

Intervention

Unitcheck provides immediate, fully automated, personalized feedback on alcohol consumption and social norms information. This feedback was available every time participants visited the website and completed the online assessment. Unitcheck was available to those in the intervention arm from weeks 1 to week 15. (An example of feedback offered and how feedback content differed from Bewick [27] can be found in Multimedia Appendix 1). The online personalized feedback consisted of three main sections:

(1) Feedback on level of alcohol consumption: Participants were presented with statements indicating the number of alcohol units they consumed per week and the associated level of health risk. Statements were standardized for each risk level (within recommended, hazardous, harmful), and gave advice about whether personal alcohol consumption should be reduced or maintained within the current sensible levels. The number of alcohol-free days was indicated, alongside information stating that it is advisable to have at least two per week. Statements related to binge drinking behavior (ie, drinking at least twice the recommended daily limit in one session) were also presented.

(2) Social norms information: Personalized statements were presented that indicated to participants the percentage of students who report drinking less alcohol than them. This was calculated relative to the risk level generated in section 1 of the feedback, and the frequency of students within each risk level was taken from data collected as part of an earlier university wide survey investigating aspects of student life in Leeds [36]. Information was also provided about the negative effects of alcohol intake reported by students who consume similar amounts of alcohol (ie, who are within the same risk category).

(3) Generic information: standard advice was provided on calculating units, the general health risks of high levels of consumption, and outlined sensible drinking guidelines publicized in the United Kingdom. Tips for sensible drinking and the contact details of both local and national support services were also presented.

Data Analysis

Previous research has suggested differential attrition according to treatment arm, and some trials have observed relatively high rates of attrition. These trial characteristics render the traditional repeated measures MANCOVA problematic, specifically liable to dropout bias. Therefore an analysis of the primary outcome data was planned that could accommodate these characteristics [37]. In order to assess the effectiveness of the intervention, the

primary outcome variable was units consumed over the past week. The data were modeled using a multilevel longitudinal regression model with time points clustered within students. That is, regression of the natural logarithm of the number of units plus 1 regressed upon male sex, assigned to intervention, age, total CAGE score, number of visits to the intervention website, and risk-taking behavior. The model was fitted on a log scale, and we took the exponential to present results on the original scale of units. It was possible that any observed effect of intervention could have been artificially produced by differential dropout rates, eg, heavier drinkers may have been less likely to complete assessments. Therefore, a logistic regression model was fitted to predict who would not complete the study. Age, units consumed the previous week at T0, sex, and treatment arm were included in the regression model. Specifically, multiple imputation was not undertaken since it depends upon the assumption that data are Missing At Random (MAR)—considered not to be likely in this situation.

Descriptive means and standard deviations were calculated for the CAGE total score, units of alcohol consumed per week and per occasion at T1, T2, and T3. Regression analysis was carried using Stata version 11.0, and descriptive statistics were carried out using SPSS v15. The data for units per week and per occasion were positively skewed, and the data were transformed before analysis was conducted. The means and standard deviations reported in the text and tables are based on untransformed data.

Results

Of the 1618 students randomly allocated, 1124 (69%) were female. Participants' age ranged from 17-50 years (mean years 20.8, SD 3.2). The majority of participants (87%) were undergraduate students, and 84% were white/white-British, based on self-reported choice from among several categories of ethnicity. The majority of the sample were UK (85%), full-time (97%) students. All 1618 students were invited to participate in the current trial. The current analysis reports on the n=1478 participants who accepted the invitation. The corresponding figures for the demographics of those who provided demographic data and are included in the current analysis are: n=1036 (70% of 1478) female, age range 17-50 years old, n=1279 (88% of 1453) white/white British, n=1282 (88% of 1459) UK student, n=1438 (99% of 1459) full-time students. Table 1 summarizes these demographics by treatment arm allocation.

Alcohol Consumption and Behavior

Of 1478 participants, 50% (n=737) reported consuming alcohol within UK government recommended weekly guidelines, 38% (n=556) at hazardous levels, and 13% (n=185) at harmful levels. Students reported consuming on average 12.7 units per occasion (SD 10.8) and 21.1 units over the last week (SD 20.9). See Table 2 for consumption by treatment arm allocation.

Table 1. Demographics of participants at baseline by treatment arm allocation (number of participants who provided demographic data is provided underneath demographic variable; percentages calculated as a percentage out of participants who provided variable data).

	Control n=755	Intervention n=723	Total n=1478
Female, n (%) n=1478	543 (71.9)	493 (68.2)	1036 (70.1)
Age, mean (SD) n=1454	20.8 (3.50)	20.8 (3.09)	20.8 (3.30)
Undergraduate, n (%) n=1459	666 (88.2)	626 (86.6)	1292 (88.6)
Full-time, n (%) n=1459	733 (98.5)	705 (97.5)	1438 (98.6)
UK student, n (%) n=1459	664 (89.2)	618 (85.5)	1282 (87.9)
White/white British, n (%) n=1453	658 (88.7)	621 (87.3)	1279 (88.0)

Table 2. Units per occasion, per previous week, and CAGE total score by treatment arm.

Consumption	Time 0		Time 1		Time 2		Time 3	
	n	M (SD)	n	M (SD)	n	M (SD)	n	M (SD)
Units consumed over the previous week [a]								
Control	755	21.7 (20.9)	544	18.0 (18.5)	380	16.3 (17.5)	321	17.1 (16.5)
Intervention	723	20.6 (20.9)	457	16.2 (16.2)	325	13.7 (15.0)	281	16.5 (18.4)
Units consumed on average drinking occasion [a]								
Control	741	12.7 (9.75)	544	10.64 (7.26)	380	10.70 (6.67)	321	9.50 (5.49)
Intervention	711	12.7 (11.8)	457	9.82 (7.13)	325	8.36 (6.21)	281	8.44 (4.87)
CAGE total score								
Control			539	1.91 (1.19)	377	1.88 (1.23)	316	1.78 (1.22)
Intervention			436	1.87(1.23)	295	1.751 (1.28)	272	1.75 (1.27)

[a]This table presents untransformed data while analysis was carried out on transformed data.

XSL•FO
RenderX

Regarding negative consequences and risk-taking behavior as a result of drinking within the past year: 34% (n=333) had injured themselves accidentally, 27% (n=248) had been injured as a result of someone else's drinking, 22% (n=195) had sexual intercourse when they ordinarily would not, 10% (n=93) had damaged property, and 3% (n=30) had caused harm to self.

Effectiveness of the Personalized Feedback and Social Norms Intervention

The variables included in the longitudinal regression model were assessment of units consumed over the last week at T1, T2, and T3; treatment arm allocation; sex; age (in years); and number of visits to intervention website. Total CAGE score, units consumed on an average drinking occasion, and reported risk taking were excluded from the final model as they did not add significantly to the model fit. The longitudinal regression model showed a significant effect of completing assessment (without intervention) on change across time with the assessment effect being greatest for those who completed T3 assessment. The model also predicted an additional effect of being assigned to intervention arm, being female, being older, and repeat visits to the intervention website.

Table 3 provides details of the regression coefficients fitted in the longitudinal model. In addition an intercept term of 3.58 corresponded to the outcome, log (1 +units consumed). It should be noted that the model identifies a lack of balance between control and intervention group at T0; the intervention group had fewer heavy drinkers. This imbalance is seen despite the stratification by unit consumption detailed in the method and despite raw observed mean values of last week consumption being similar between arms (see Discussion for further comment; see Table 2). The model yielded an overall R^2 value of 0.05 and an interclass correlation coefficient of .24, indicating that there was significant variation between participants and over time. The transformation makes the model hard to interpret directly, and so we have calculated examples in Table 4. For example, the model predicted that a typical 21-year-old female allocated to control who completed T1 assessment would, at week 34, drink 13.33 units per week while the corresponding figure for males was 19.89 units. As can be seen in Table 4, when students completed T3 assessment, consumption decreased to 12.43 for females and 18.54 for males. When assigned to the intervention arm, there was an additional effect with the model predicting that at week 34 females in the intervention condition would drink 9.49 units per week while males would drink 14.15 units. There was an additional effect of multiple visits to the intervention website. The model predicted females who visited the site three times would drink 5.87 units per week while males would drink 8.76 units. Despite the variation in individual drinking patterns across time, the data included enough observations to see an effect of the intervention.

Regarding adherence, a typical participant completed between two and three of the four assessments (mean assessments 2.6). The logistic regression model showed that the risk of dropping out after baseline was increased by being assigned to the intervention and drinking more at baseline; neither sex, age, nor total CAGE score added significantly to the model once these variables were taken into account. After completing T1 assessment, there was no clear pattern concerning dropout—attrition appeared to be random and not predicted by any of the covariates recorded.

Table 3. Table of coefficients for longitudinal regression model: log (1+units consumed over the last week) regression on assessment completed, condition allocation, sex, age, and number of visits to website by restricted maximum likelihood.

Covariate	Coefficient	95% CI	P value
Complete assessment T1	-.15	-0.25 to -0.06	.001
Complete assessment T2	-.36	-0.47 to -0.25	<.001
Complete assessment T3	-.24	-0.35 to -0.13	<.001
Allocated to receive feedback	-.27	-0.41 to -0.13	<.001
Male	.40	0.32 to 0.48	<.001
Age	-.04	-0.05 to -0.03	<.001
Number of visits to feedback website	-.16	-.21 to -0.11	<.001
Constant	3.58	3.32 to 3.84	<.001

Table 4. Prediction of units consumed over the last week at each time point (longitudinal regression model).

	Female 21 years old	Male 21 years old					Female 21 years old	Male 21 years old			
	Allocated to control	Allocated to intervention					Allocated to control	Allocated to intervention			
# of visits to intervention		0	1	2	3	4		0	1	2	3	4
Completed assessment at T0	15.49	11.82					23.10	17.64				
Completed assessment at T1	13.33	10.18					19.89	15.18				
Completed assessment at T2	10.80	8.25	7.03	5.99	5.10	4.35	16.12	12.30	10.49	8.94	7.61	6.46
Completed assessment at T3	12.43	9.49	8.08	6.89	5.87	5.00	18.54	14.15	12.06	10.28	8.76	7.46

Discussion

Principal Findings

This study aimed to evaluate the effectiveness of Unitcheck. The model predicted a monitoring effect, with participants who completed assessments reducing alcohol consumption over the last week. Further reductions were predicted for those allocated to receive the intervention, and additional reductions were predicted as the number of visits to the intervention website increased. The model therefore supported the hypothesis that Unitcheck, a Web-based social norms intervention, can reduce the amount of alcohol consumed over the last week. The model did not predict a reduction of units consumed on an average occasion. The results also suggest that the reduction can be sustained in the medium-term (ie, 19 weeks after access to the intervention was closed).

The previous feasibility trial reported significant reductions in units consumed per occasion but not in units consumed over the last week [27]. In this replication study, assessment of units consumed over the last week was carried out by providing participants with a list of common alcoholic beverages and asking them to indicate how many they had consumed over the last 7 days. In the current trial, the assessment was altered; participants were provided with a list of common alcohol beverages and were asked to indicate how many they had consumed on each day over the last 7 days (ie, 7-day recall). The current sample reported higher levels of consumption when compared to the feasibility sample. It is unclear whether this difference is due to differences in recording or actual behavior.

The current study findings are consistent with our multisite trial [12] that observed an effect of assessment across time on units consumed in the previous week; an additional effect of being assigned to receive the intervention was also predicted. The current study predicted a monitoring effect, and the multisite study results supports this finding with the greatest reductions being observed among participants who were monitored (ie,

completed at least 2 of the 5 assessments). In both studies, there was an additional effect of being allocated to the intervention arm.

It is a strength of the current study that participants reported a range of levels of consumption (from within sensible guidelines to hazardous drinking). Unitcheck was designed as a public health intervention that could be delivered across the student population. In contrast, previous studies have reported a large proportion of low-level consumers [28], limiting the potential to see any significant decrease in consumption.

Since, after T1, dropping out is not related to previous drinking behavior, the changes in drinking are not due to completers being the lighter drinkers; this is a further strength of the study. Prior to completing T1, the risk of dropping out was increased by being assigned to the intervention and drinking more at baseline. This is consistent with previous research report of higher levels of attrition among heavier consumers of alcohol [27,38]. This suggests further work is needed to consistently engage students who are currently consuming alcohol at potentially problematic levels. In addition, it is necessary that we understand the processes by which participants choose to engage with research investigating Web-based interventions and, ultimately, how to encourage increased levels of engagement with interventions.

A common method used to investigate the influence of dropout from longitudinal studies is multiple imputation. Multiple imputation is dependent on the assumption that data are MAR. In the current study, we consider MAR unlikely; therefore, multiple imputation was not used in the analysis.

Limitations

This RCT included a medium-term postintervention follow-up. This, combined with the relatively large numbers of participants recruited and retained (compared with previous studies in this area [22,27]), means it makes a distinctive contribution to the evidence base. However, a number of limitations need to be

considered when interpreting the results. First, the intervention group had fewer heavy drinkers. This does not necessarily detract from the findings reported but is an issue for concern. The attempt to stratify by four confounders was too ambitious. As a consequence, the stratification by alcohol units was too crude and the imbalance occurred. Second, the study design randomized individuals after registering interest but before providing full baseline assessment. This meant that 71% of those randomized accepted the invitation to participate and provided T1 assessment. Third, although 74% of intervention participants accessed the intervention, the proportion who engaged with follow-up assessments was lower (with 43% of intervention participants completing all assessments; 47% of control participants). High dropout is a concern since it might explain the findings rather than the monitoring or intervention. For example, if heavier drinkers drop out, then the average level of drinking of those retained will decline over time. To explore this, we investigated models for dropout. There was evidence of an association between heavier drinking and dropout after T0 but not beyond that time. We note also that at T3, the average level of drinking increases rather than decreases; this is inconsistent with the "alternative" but consistent with effects of monitoring and intervention wearing off over time. Fourth, while there was a 34-week follow-up assessment, these results say little about the longer-term impact of the intervention. The longevity of electronic brief interventions is still uncertain, but the current results suggest that repeated access to such interventions might help maintain behavior change. Fifth, participants were not blind to their condition as participants were aware of whether or not they received feedback. Control participants were aware that at the end of the study they would gain access to personalized feedback. Sixth, there were two small differences in the treatment of the intervention and control groups (intervention participants could receive up to £0.25 more than control participants; intervention participants received an extra email contact reminding them to visit the website).

Conclusions

These results lend further support to the efficacy and potential effectiveness of using Web-based interventions to reduce alcohol consumption among the student population. The findings add weight to the suggestion that one active ingredient to Web-based personalized feedback is the self-monitoring support they afford to individuals. By adding a postintervention follow-up, this study supports the idea that behavior change instigated as a result of engaging with Web-based interventions can be sustained, at least in the short- to medium-term. Future research should seek to investigate the generalizability of these findings to other sections of the general population. In addition, further work is needed to understand the mechanisms of engagement and behavior change, in the hope of further enhancing the impact of brief Web-based interventions.

Conflicts of Interest

In the past, Bewick, as keynote speaker, has received reimbursement of travel expenses from Anheuser-Busch and Noctis.

Multimedia Appendix 1

Original feedback with details of how feedback was altered between the feasibility study (Bewick et al, 2008) and the current study.

[PDF File (Adobe PDF File), 350KB - jmir_v15i7e137_app1.pdf]

Multimedia Appendix 2

CONSORT-EHEALTH checklist V1.6.2 [39].

[PDF File (Adobe PDF File), 350KB - jmir_v15i7e137_app1.pdf]

References

1. Gill JS. Reported levels of alcohol consumption and binge drinking within the UK undergraduate student population over the last 25 years. Alcohol and Alcoholism 2002;37(2):109-120. [doi: 10.1093/alcalc/37.2.109]
2. Schulenberg JE, Maggs JL. A developmental perspective on alcohol use and heavy drinking during adolescence and the transition to young adulthood. J Stud Alcohol Suppl 2002;14:54-70. [Medline: 12022730]
3. Royal College of Psychiatrists. Mental health of students in higher education. College Report 2011:166.
4. White HR, Morgan TJ, Pugh LA, Celinska K, Labouvie EW, Pandina RJ. Evaluating two brief substance-use interventions for mandated college students. J Stud Alcohol 2006 Mar;67(2):309-317. [Medline: 16562414]
5. Delk EW, Meilman PW. Alcohol use among college students in Scotland compared with norms from the United States. J Am Cll Health 1996;44(6):274-281. [Medline: 8735165]

XSL•FO
RenderX

6. Orford J, Waller S, Peto J. Drinking behavior and attitudes and their correlates among university students in England. I. Principal components in the drinking domain. II. Personality and social influence. III. Sex differences. Q J Stud Alcohol 1974 Dec;35(4 Pt A):1316-1374. [Medline: 4445468]

7. Wechsler H, Davenport A, Dowdall G, Moeykens B, Castillo S. Health and behavioural consequences of binge drinking in college. A national survey of students at 140 campuses. JAMA 1994;272(21):1672-1677. [Medline: 7966895]

8. Wechsler H, Lee JE, Kuo M, Seibring M, Nelson TF Lee H. Trends in college binge drinking during a period of increased prevention efforts: Findings from 4 Harvard School of Public Health College Alcohol Study surveys: 1993-2001. Journal of American College Health 2002;50:203-217. [Medline: 11990979]

9. Smit F, Lokkerbol J, Riper H, Majo MC, Boon B, Blankers M. Modeling the cost-effectiveness of health care systems for alcohol use disorders: how implementation of eHealth interventions improves cost-effectiveness. J Med Internet Res 2011;13(3):e56 [FREE Full text] [doi: 10.2196/jmir.1694] [Medline: 21840836]

10. Blankers M, Nabitz U, Smit F, Koeter MWJ, Schippers GM. Economic evaluation of internet-based interventions for harmful alcohol use alongside a pragmatic randomized controlled trial. J Med Internet Res 2012;14(5):e134 [FREE Full text] [doi: 10.2196/jmir.2052] [Medline: 23103771]

11. Kypri K, Hallett J, Howat P, McManus A, Maycock B, Bowe S, et al. Randomized controlled trial of proactive web-based alcohol screening and brief intervention for university students. Arch Intern Med 2009 Sep 14;169(16):1508-1514. [doi: 10.1001/archinternmed.2009.249] [Medline: 19752409]

12. Bewick BM, West R, Gill J, O'May F, Mulhern B, Barkham M, et al. Providing web-based feedback and social norms information to reduce student alcohol intake: a multisite investigation. J Med Internet Res 2010;12(5):e59 [FREE Full text] [doi: 10.2196/jmir.1461] [Medline: 21169171]

13. Cunningham JA, Wild TC, Cordingley J, van Mierlo T, Humphreys K. A randomized controlled trial of an internet-based intervention for alcohol abusers. Addiction 2009;104(12):2023-2023. [Medline: 19922569]

14. Doumas DM, McKinley LL, Book P. Evaluation of two Web-based alcohol interventions for mandated college students. J Subst Abuse Treat 2009 Jan;36(1):65-74. [doi: 10.1016/j.jsat.2008.05.009] [Medline: 18657941]

15. Foxcroft D, Ireland D, Lister-Sharp DJ, Lowe G, Breen R. Longer-term primary prevention for alcohol misuse in younger people: A systematic review. Addiction 2003;98:397-411. [Medline: 12653810]

16. McAlaney J, Bewick B, Hughes C. The international development of the 'Social Norms' approach to drug education and prevention. Drugs: education, prevention and policy 2011;118(2):81-89.

17. Clapp JD, McDonnell AL. The relationship of perceptions of alcohol promotion and peer drinking norms to alcohol problems reported by college students. Journal of College Student Development 2000;41:20-26.

18. Perkins HW, Wechsler H. Variation in perceived college drinking norms and its impact on alcohol abuse: A nationwide survey. Journal of Drug Issues 1996;26:961-974.

19. Moreira MT, Smith LA, Foxcroft D. Social norms interventions to reduce alcohol misuse in University or College students. Cochrane Database of Systematic Reviews 2009;3:---. [doi: 10.1002/14651858.CD006748.pub2]

20. Bewick BM, Trusler K, Barkham M, Hill AJ, Cahill J, Mulhern B. The effectiveness of web-based interventions designed to decrease alcohol consumption--a systematic review. Prev Med 2008 Jul;47(1):17-26. [doi: 10.1016/j.ypmed.2008.01.005] [Medline: 18302970]

21. Elliot JC, Carey KB, Bolles JR. Computer-based interventions for college drinking: A qualitative review. Addictive Behaviors 2008;33:994-1005. [doi: 10.1016/j.addbeh.2008.03.006]

22. Khadjesari Z, Murray E, Hewitt C, Hartley S, Godfrey C. Can stand-alone computer-based interventions reduce alcohol consumption? A systematic review. Addiction 2010;106:267-282. [Medline: 21083832]

23. Moreira MT, Oskrochi R, Foxcroft DR. Personalised normative feedback for preventing alcohol misuse in university students: Solomon three-group randomised controlled trial. PLoS One 2012 Sep;7(9):e44120 [FREE Full text] [doi: 10.1371/journal.pone.0044120] [Medline: 22984466]

24. Riper H, van Straten A, Keuken M, Smit F, Schippers G, Cuijpers P. Curbing problem drinking with personalized-feedback interventions: a meta-analysis. Am J Prev Med 2009 Mar;36(3):247-255. [doi: 10.1016/j.amepre.2008.10.016] [Medline: 19215850]

25. Walters ST, Vader AM, Harris TR. A Controlled Trial of Web-based Feedback for Heavy Drinking College Students. Prevention Science 2007;8(2):83-88. [Medline: 17136461]

26. Doumas DM, Anderson L. Reducing alcohol use in first-year university students: Evaluation of a web-based personalized feedback program. Journal of College Counseling 2009;12(1):18-32. [doi: 10.1002/j.2161-1882.2009.tb00037.x]

27. Bewick BM, Trusler K, Mulhern B, Barkham M, Hill AJ. The feasibility and effectiveness of a web-based personalised feedback and social norms alcohol intervention in UK university students: A randomised control trial. Addictive Behaviors 2008;33:1192-1198. [Medline: 18554819]

28. Kypri K, McAnally HM. Randomized controlled trial of a web-based primary care intervention for multiple health risk behaviors. Prev Med 2005;41(3-4):761-766. [doi: 10.1016/j.ypmed.2005.07.010] [Medline: 16120456]

29. Kypri K, Langley JD, Saunders JB, Cashell-Smith ML, Herbison P. Randomized controlled trial of web-based alcohol screening and brief intervention in primary car. Archives Internal Medicine 2008;168(5):530-536. [Medline: 18332300]

30. Ewing JA. Detecting alcoholism. The CAGE questionnaire. JAMA 1984 Oct 12;252(14):1905-1907. [Medline: 6471323]

XSL•FO

RenderX

31.   Mayfield D, McLeod G, Hall P. The CAGE questionnaire: validation of a new alcoholism screening instrument. Am J Psychiatry 1974 Oct;131(10):1121-1123. [Medline: 4416585]

32.   Aertgeerts B, Buntinx F, Bande-Knops C, Vandermeulen M, Roelants M, Ansoms S, et al. The value of CAGE, CUGE, and AUDIT in screening for alcohol abuse and dependence among college freshmen. Alcohol Clin Exp Res 2000 Jan;24(1):53-57. [Medline: 10656193]

33.   Shields AL, Caruso JC. A Reliability Induction and Reliability Generalization Study of the Cage Questionnaire. Educational and Psychological Measurement 2004 Apr 01;64(2):254-270. [doi: 10.1177/0013164403261814]

34.   Dawson DA. Methodological issues in measuring alcohol use. Alcohol Research and Health 2003;27(1):18-29. [Medline: 15301397]

35.   Department of Health. The next steps in the National Alcohol Strategy. London, UK; 2007. Safe. Sensible. Social URL: http://news.bbc.co.uk/2/shared/bsp/hi/pdfs/5_6_07_alcoholsurvey.pdf [accessed 2013-06-12] [WebCite Cache ID 6HJuZ7y8n]

36.   Bewick BM, Mulhern B, Barkham M, Trusler K, Hill AJ, Stiles WB. Changes in undergraduate student alcohol consumption as they progress through university. BMC Public Health 2008;8:163 [FREE Full text] [doi: 10.1186/1471-2458-8-163] [Medline: 18489734]

37.   Laird NM, Ware JH. Random-effects models for longitudinal data. Biometrics 1982;38(4):963-974.

38.   Edwards A, Rollnick S. Outcome studies of brief alcohol intervention in general practice: the problem of lost subjects. Addiction 1997 Dec;92(12):1699-1704. [Medline: 9581002]

39.   Eysenbach G, CONSORT-EHEALTH Group. 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]

Abbreviations

CAGE: (1) have you ever thought about Cutting down on your drinking, (2) do you ever get Annoyed at criticism of your drinking, (3) do you ever feel Guilty about your drinking, and (4) do you ever have a drink in the morning (an Eye-opener)

MANCOVA: multivariate analysis of covariance

MAR: missing at random

RCT: randomized controlled trial

Original Paper

Maximizing the Value of Mobile Health Monitoring by Avoiding Redundant Patient Reports: Prediction of Depression-Related Symptoms and Adherence Problems in Automated Health Assessment Services

John D Piette[1], PhD, ScM; Jeremy B Sussman[1], MD; Paul N Pfeiffer[2], MD; Maria J Silveira[1], MD; Satinder Singh[3], PhD; Mariel S Lavieri[4], PhD

[1]VA Center for Clinical Management Research and Division of General Medicine, Department of Internal Medicine, University of Michigan, Ann Arbor, MI, United States

[2]VA Center for Clinical Management Research and Department of Psychiatry, Ann Arbor VA Healthcare System and University of Michigan, Ann Arbor, MI, United States

[3]Artificial Intelligence Laboratory, Department of Electrical Engineering and Computer Science, College of Engineering, University of Michigan, Ann Arbor, MI, United States

[4]Deparment of Industrial and Operations Engineering, College of Engineering, University of Michigan, Ann Arbor, MI, United States

Corresponding Author:
John D Piette, PhD, ScM
VA Center for Clinical Management Research and Division of General Medicine
Department of Internal Medicine
University of Michigan
PO Box 130170
Ann Arbor, MI, 48113-0170
United States
Phone: 1 734 936 4787
Fax: 1 734 936 8944
Email: jpiette@umich.edu

Abstract

Background: Interactive voice response (IVR) calls enhance health systems' ability to identify health risk factors, thereby enabling targeted clinical follow-up. However, redundant assessments may increase patient dropout and represent a lost opportunity to collect more clinically useful data.

Objective: We determined the extent to which previous IVR assessments predicted subsequent responses among patients with depression diagnoses, potentially obviating the need to repeatedly collect the same information. We also evaluated whether frequent (ie, weekly) IVR assessment attempts were significantly more predictive of patients' subsequent reports than information collected biweekly or monthly.

Methods: Using data from 1050 IVR assessments for 208 patients with depression diagnoses, we examined the predictability of four IVR-reported outcomes: moderate/severe depressive symptoms (score ≥10 on the PHQ-9), fair/poor general health, poor antidepressant adherence, and days in bed due to poor mental health. We used logistic models with training and test samples to predict patients' IVR responses based on their five most recent weekly, biweekly, and monthly assessment attempts. The marginal benefit of more frequent assessments was evaluated based on Receiver Operator Characteristic (ROC) curves and statistical comparisons of the area under the curves (AUC).

Results: Patients' reports about their depressive symptoms and perceived health status were highly predictable based on prior assessment responses. For models predicting moderate/severe depression, the AUC was 0.91 (95% CI 0.89-0.93) when assuming weekly assessment attempts and only slightly less when assuming biweekly assessments (AUC: 0.89; CI 0.87-0.91) or monthly attempts (AUC: 0.89; CI 0.86-0.91). The AUC for models predicting reports of fair/poor health status was similar when weekly assessments were compared with those occurring biweekly (P value for the difference=.11) or monthly (P=.81). Reports of medication adherence problems and days in bed were somewhat less predictable but also showed small differences between assessments attempted weekly, biweekly, and monthly.

Conclusions: The technical feasibility of gathering high frequency health data via IVR may in some instances exceed the clinical benefit of doing so. Predictive analytics could make data gathering more efficient with negligible loss in effectiveness. In particular, weekly or biweekly depressive symptom reports may provide little marginal information regarding how the person is doing relative to collecting that information monthly. The next generation of automated health assessment services should use data mining techniques to avoid redundant assessments and should gather data at the frequency that maximizes the value of the information collected.

Introduction

Clinicians and health care payers increasingly look to mobile health services such as Interactive Voice Response (IVR) as tools for monitoring patients' status between face-to-face encounters and identifying individuals who need attention to prevent acute events [1-3]. Multiple studies have shown that IVR monitoring yields actionable and reliable clinical information even on sensitive topics such as mental health and substance abuse [4-11]. Moreover, patients are willing to complete regular IVR assessments over extended periods of time, even when challenged by chronic illness, age, poverty, low literacy, and psychiatric problems [12,13].

While IVR has significant potential to increase the information base of proactive care management, the design of automated monitoring services can have negative consequences that should be carefully considered when deciding the frequency and content of each assessment call. Studies suggest that patients may tire of frequent IVR assessments [12-15], particularly if they are asked repeatedly for information about health or self-care problems that have not changed. At the same time, many patients have a large number of health problems associated with multiple chronic conditions [16,17]. For such patients, current alternatives to the typical disease-specific focus include substantially increasing the length of each assessment, increasing the frequency of assessment calls, focusing on a broader number of problems but with less depth on each, or focusing only on cross-cutting issues such as medication adherence or physical activity. Each of these strategies introduces new challenges to sustaining patient engagement or the quality of information for clinical decisions. As with other types of patient contact [18-21], the timing and content of IVR monitoring is almost always based on expert opinion and static flow diagrams. As such, these systems have not achieved their full potential as a strategy for cost-effectively increasing patients' access to between-visit monitoring and self-care support.

While frequent (eg, weekly or daily) IVR assessment calls may be necessary to detect fluctuations in important health indicators, what if a patient's IVR assessment reports could be predicted based on the information that he or she provided in prior calls? For example, if a patient has consistently reported perfect medication adherence over multiple prior IVR assessments, what would be the probability that they would report something different today? Data mining is a set of analytic techniques designed to extract latent information from data in order to make predictions about the future [22,23]. In the context of IVR, data mining could help identify when patients' answers are so stable that the same questions are not worth asking again, or when there are changes in the patient's status indicating the need for more intensive probing. Using information about such patterns, adaptive mobile health monitoring programs could be developed that automatically adjust the frequency and content of assessments so that they provide the most useful information for guiding patient counseling and clinical follow-up.

We used one approach to data mining in order to examine data from 1050 IVR assessments of 208 patients with depression diagnoses. All patients received IVR calls at regular intervals, during which they completed the Patient Health Questionnaire (PHQ-9) [24,25], a widely used and validated depression assessment scale. Also, patients repeatedly answered questions regarding their antidepressant medication adherence, perceived general health, and days in bed due mental health problems. Given the large number of serial reports from each patient, we examined the predictability of patients' IVR responses. Specifically, for each patient we identified the five most recent weekly, biweekly, and monthly assessments. We used those data plus other information collected during prior assessments and at the time of the patient's enrollment to determine the extent to which health reports were predictable and whether that predictability varied according to the frequency of attempted assessment calls. Based on these analyses, we determined whether less frequently collected data (eg, biweekly or monthly) provided as much information about patients' status as information collected weekly, thereby making it possible to decrease the frequency of IVR calls or to change their focus to other important health indicators. More generally, we sought to determine whether data mining techniques might inform automated assessments that repeatedly measure patients' health status, so that the most clinically useful, nonredundant information is collected.

Methods

Patient Eligibility and Recruitment

Patients were enrolled between March 2010 and January 2012 from 13 university-affiliated and community-based primary care practices. To be eligible, patients had to have two primary care visits in the previous 2 years, at least one in the previous 13 months, and either a depression diagnosis listed in clinical records or an antidepressant prescription plus a diagnosis of depression listed in billing data. Patients with schizophrenia, psychosis, delusional disorder, bipolar disorder, or dementia were excluded. Potential participants were mailed an

introductory letter that was followed by a screening and recruitment telephone call. Patients who provided informed consent were enrolled in the IVR system and mailed additional program information, including materials describing effective communication with informal caregivers and clinicians. The study was approved by the human subjects committees of the University of Michigan and Ann Arbor VA Healthcare System. More information about the intervention and patients' engagement in the IVR calls has been published elsewhere [13].

IVR Monitoring Protocol

Detailed information about the IVR call content and functioning are available by contacting the authors. In brief, each week that an assessment was scheduled, the system made up to three attempts to contact the patient on up to three different patient-selected day/time combinations. The content of the calls was developed with input from psychiatrists, primary care providers, and experts in IVR program design and health behavior change. Every call included an assessment of patients' depression symptoms using the PHQ-9 [24]. The PHQ-9 is a 9-item questionnaire that is sensitive and specific with respect to other established measures of major depression. Scores are associated with physical functioning, sick days, and health care use [24]. Because self-rated health status is correlated with patients' service use and mortality risk [26-28], they were asked the standard item, "Thinking about your overall health, how were you feeling this past week (excellent, very good, good, fair, poor)?" Medication adherence was assessed by asking: "How often during the past week did you take your depression medication exactly as prescribed (always, most of the time, less than half of the time, rarely or never)?" Finally, during each assessment, patients were asked: "This past week, did you ever stay in bed all or most of the day because of your mental health (yes versus no)?" Calls used tree-structured algorithms to present recorded queries and tailored information that was invariant across patients and over time. Based on patients' responses, they received tailored advice for managing their self-care. For example, patients' received messages tailored according to their recent trajectories in depression scores (trending positive, negative, or stable and by how much), including messages such as the following:

> It sounds like you're still experiencing some serious symptoms of depression. Remember that if you're prescribed a medication for depression, it's important that you keep taking it exactly as prescribed to keep your depression from getting worse. Sometimes it takes awhile for a depression medication to work, so if you have been on your current medication for less than 8 weeks, try to be patient and see if you start to see some improvement. If you've been on the same medication for more than 8 weeks and you're still not feeling okay, your doctor wants to know. You should make an appointment with your doctor to talk about some other treatment options. I'll give you the phone number of your doctor's office at the end of this call.

Clinicians received fax alerts identifying patients reporting health problems requiring follow-up before their next outpatient encounter. For patients enrolling with a family caregiver, those caregivers received automatic updates by IVR and email with suggestions regarding how they could support the patient's self-management.

Outcomes of Interest

For each assessment, we created binary indicators for each of the four outcomes reported: (1) moderate/severe depressive symptoms indicated by a PHQ-9 score of ≥10; (2) fair or poor perceived general health status; (3) poor antidepressant adherence, ie, rarely or never taking antidepressant medication as prescribed; and (4) spending days in bed in the past week due to mental health problems.

Analytic Sample Definition and Analyses

In order to determine the predictability of patients' assessment reports based on the content and frequency of prior assessments, we identified the subset of patients with one or more "index" assessments meeting the following criteria: (a) five completed prior assessments immediately preceding the index assessment and collected with the program's normal frequency of weekly assessment attempts; (b) five completed prior assessments with a 2-week minimum gap between each one; and (c) five completed prior assessments with a minimum 4-week gap between each one. A total of 1050 index assessments for 208 unique patients were identified.

In addition to linking each index assessment to prior assessment information, index assessments also were linked with information about that patient's sociodemographic and clinical characteristics collected at the time of program initiation. Those baseline data included patients' age, gender, educational attainment, baseline depressive symptom severity score (ie, measured using the PHQ-9 minus the item asking about suicidal ideation [29]), self-reported hospital admission in the year prior to program entry, physical functioning as measured by the SF-12 [30], and the number of comorbid chronic medication conditions.

In initial analyses, we examined the correlation across the four health indicators reported within each index assessment, and we calculated the alpha reliability of patients' IVR-reported PHQ-9 scores. We then examined the proportion of patients reporting each health problem in the index assessment when the same problem was reported in the one or in both of the most recent prior assessments assuming weekly, biweekly, or monthly assessment attempts. For example, we examined measures of association between patient reports of moderate/severe depressive symptoms (PHQ-9 ≥10) and similarly high PHQ-9 scores in the most recent assessment or both of the two most recent assessments (assuming weekly, biweekly, and monthly assessment calls).

Finally, we fit multivariate logistic regression models predicting each of the four health indicators as reported in index assessments. Each model included patients' baseline sociodemographic and clinical characteristics as defined above, as well as information about that same health indicator and the other three health indicators reported in five prior assessments collected assuming a periodicity of weekly, biweekly, or monthly call attempts. Serial indicators designed to capture additional information about trends in patients' depression scores

(eg, the number of weeks since program entry and prior number of completed assessments) also were considered as potential predictors. For models predicting moderate/severe depressive symptoms, fair/poor health, and days in bed, these additional variables had no discernible marginal predictive value in the context of the multiple prior, ordered indicators of the patient's health and self-care. However, an indicator for weeks since program entry was a marginally significant predictor of patients' medication adherence and was retained in the models used as the basis of ROC curves predicting patient reports of poor antidepressant medication adherence.

When fitting each of the three models, we used two strategies to prevent overfitting to the current dataset. First, we used 10-fold cross validation, in which the model was fit 10 times based on random 90% training samples and then used to predict the outcomes in mutually exclusive 10% test samples. Second, for each of the ten replications, we used stepwise regression (with a *P* value of .20 for removal) to identify the most significant subset of candidate predictors. All models also adjusted for clustering of assessment responses by patient.

The predictive significance of the three models for each outcome was compared graphically to one another and to a model with only baseline information using Receiver Operator Characteristic (ROC) curves. We also compared the area under the curve (AUC) across ROCs and calculated each AUC's 95% confidence interval [31]. To illustrate the potential predictive accuracy of the best model for each outcome, we report the sensitivity and specificity at the point on the ROC curve with the highest proportion of outcomes correctly predicted.

Results

Patient Characteristics

Patients were on average 52.2 years of age. Most were women, white, and married (Table 1). Patients reported a mean of 2.4 comorbid chronic conditions including hypertension (50.0%), arthritis (49.5%), chronic lung disease (33.2%) and back pain (42.1%). Roughly a third (33.2%) of patients had moderate or severe depressive symptoms at baseline; those patients were somewhat younger on average at the time of program enrollment than patients with mild depressive symptoms.

Co-Occurrence of Reported Health Problems Within IVR Assessments

Patients reporting a given problem during their IVR assessments were more likely to report other concurrent problems as well. For example, compared to patients reporting mild depressive symptoms, those reporting moderate/severe depressive symptoms were more likely also to report staying in bed all or most of the day due to mental health problems (27% versus 8%) and that their general health was either fair or poor (47% versus 14%, both *P*<.001 after adjusting for clustering by patient). Similarly, patients reporting being bedbound during the past week due to mental health problems were significantly more likely than other patients to rate their health as fair or poor during the same assessment (29% versus 20%, *P*<.001). Patients reporting that they rarely or never took their medication as

prescribed were more likely than other patients to report poor general health (28% versus 17%; *P*<.001).

Bivariate Relationship Between IVR Reports and Prior Reports of the Same Outcome

The internal reliability of the PHQ-9 was excellent (alpha=.87). Patients were substantially more likely to report moderate/severe depressive symptoms if they reported similar information in prior assessments (Table 2). For example, while patients reported moderate/severe depressive symptoms in 21.5% of all assessments, they did so 70.3% of the time when they also reported similarly high symptoms on their most recent assessment, and 83.3% of the time when they reported moderate/severe depressive symptoms during both of their most recent assessments, assuming weekly assessment attempts. Ninety-one percent of patients whose most recent weekly PHQ-9 score was <10 also had a score <10 on their index assessment. Assuming weekly assessment attempts, a similar pattern was observed with respect to the autocorrelation of patients' reported general health status, medication adherence, and days in bed due to mental health problems.

In general, assessments collected biweekly or monthly were only somewhat less correlated with subsequent reports than information collected assuming weekly assessment attempts. For example, 58.8% of index assessments in which the patient reported moderate/severe depressive symptoms had similarly high levels in the two most recent assessments collected assuming weekly attempts, as compared to 53.4% on the two prior assessments collected biweekly, and 51% on the two prior assessments collected monthly.

Predictive Models

Moderate/Severe Depression

ROC curves for models predicting patients' depressive symptoms were highly predictive with an AUC≥0.89 regardless of whether prior assessments were attempted weekly, biweekly, or monthly (Figure 1 and Table 3). In Figure 1, the blue line represents weekly assessment attempts, the green line represents biweekly attempts, and the red line represents monthly attempts. The yellow line represents the ROC curve for the model predicting depressive symptoms using baseline data only. All other models also included baseline clinical and sociodemographic information. While the AUC for weekly assessments was significantly different than either biweekly (*P*<.001) or monthly assessments (*P*<.001), there was no statistically significant difference in the AUC for biweekly compared to monthly calls (*P*=.36).

The AUC for the model assuming weekly assessment attempts was .91 (95% CI 0.89, 0.93). At the point on the ROC curve with the greatest number of reports correctly classified (ie, a probability of moderate/severe depression=.50), 88.4% of assessments were classified correctly with a sensitivity of .68 and a specificity of .94. As expected, regardless of the frequency of assessment attempts, patients' prior PHQ-9 scores were the strongest predictor of index assessment scores ≥10, although prior IVR reports regarding general health status, baseline depressive symptom severity, baseline physical functioning,

and the number of comorbidities reported at baseline also were significant independent predictors of patients' depression status.

General Health Status

Similar to patients' reports of their depressive symptoms, reports of perceived general health status were highly predictable based on prior information (Figure 2). In Figure 2, the blue line represents weekly assessment attempts, the green line represents biweekly attempts, and the red line represents monthly assessment attempts. The yellow line represents the prediction based on baseline data only. All other models also included baseline clinical and sociodemographic information.

The AUC for the model assuming weekly assessment attempts was 0.88 (95% CI 0.86, 0.91). The AUC for that model was not statistically different from the one assuming biweekly attempts (P=.11) or assessments collected monthly (P=.81). Prior reports of perceived health status were the strongest predictors, although prior information about days in bed due to mental health problems and about medication adherence problems also were consistently retained in logistic models as predictors of patients' index assessment reports of fair/poor health. With respect to the model assuming weekly assessment attempts, the cutoff indicating a probability of fair/poor health=.50 correctly classified 87% of all index assessments, with a sensitivity of .58 and a specificity of .95.

Table 1. Patient characteristics (cell entries, aside from N, are either column percent or mean [SD]).

| | Depressive symptom severity[a] | | | |
	Total	Moderate/Severe	Mild	P value
N	208	69	139	
Age in years	52.2 (12.5)	50.6 (12.0)	53.7 (12.8)	.04
Female	79.0	77.9	80.0	.72
White	90.0	89.5	90.5	.81
Married	60.0	57.9	62.1	.55
More than high school	79.5	75.8	83.2	.21
Prior hospitalization[b]	21.6	24.2	19.0	.38
Number of diagnoses	2.4 (1.7)	2.6 (1.7)	2.2 (1.6)	.09
Hypertension	50.0	55.8	44.2	.11
Cardiovascular disease	8.4	10.5	6.3	.30
Stroke	4.2	4.2	4.2	1.00
Arthritis	49.5	52.6	46.3	.38
Chronic lung disease	33.2	41.1	25.3	.02
Back pain	42.1	43.2	41.1	.77
Physical functioning[c]	39.6 (13.8)	37.8 (14.2)	41.4 (13.3)	.07

[a]PHQ-9: 9-item Patient Health Questionnaire score ≥10 or <10.

[b]1+ hospitalizations in the year prior to enrollment.

[c]Physical Functioning: 12-item Medical Outcome Study Short Form Physical Composite Summary. Scores range from 0 to 100 with higher scores indicating greater functioning.

Table 2. Variation in problem reports by the number and frequency of prior reports of the same problem.

	Weekly		Biweekly		Monthly	
	1 report[a]	2 reports[b]	1 report	2 reports	1 report	2 reports
Moderate/Severe Depression [c]						
% with Report[d]	21.5	21.5	21.5	21.5	21.5	21.5
Sensitivity[e]	69.6	58.8	69.1	53.4	64.2	51.0
Specificity[f]	92.0	96.8	90.5	95.4	89.8	95.8
PPV[g]	70.3	83.3	66.5	76.2	63.3	77.0
NPV[h]	91.7	89.6	91.5	88.2	90.2	87.7
Fair/Poor Health						
% with Report	21.4	21.4	21.4	21.4	21.4	21.4
Sensitivity	67.2	57.4	67.2	57.4	67.7	55.9
Specificity	90.5	96.7	89.4	96.7	89.7	96.4
PPV	65.9	82.4	63.4	82.4	64.2	80.9
NPV	91.0	89.3	90.9	89.3	91.0	88.9
Poor Adherence						
% with Report	18.6	18.6	18.6	18.6	18.6	18.6
Sensitivity	55.5	43.2	54.2	42.6	58.4	39.6
Specificity	90.4	96.2	91.2	96.6	90.1	96.5
PPV	57.0	72.0	58.3	74.2	57.3	71.8
NPV	89.9	88.1	89.7	88.1	90.5	87.5
In Bed Due to Mental Health						
% with Report	12.9	12.9	12.9	12.9	12.9	12.9
Sensitivity	45.4	24.2	39.5	17.5	30.3	11.7
Specificity	91.8	97.3	91.0	97.3	90.7	97.4
PPV	45.0	55.8	39.5	47.7	32.7	38.9
NPV	91.9	90.0	91.0	89.2	89.7	88.6

[a]Patient also reported the same health problem in the most recent assessment during the time frame.

[b]Patient also reported the same health problem in the two most recent assessments during the time frame.

[c]PHQ-9 score ≥10.

[d]Percentage of all index assessments in which that health problem was reported.

[e]Proportion of index assessments reporting that health problem that also had the problem reported in the prior assessment(s).

[f]Proportion of index assessments not reporting that health problem that also were negative in the prior assessment(s).

[g]PPV: Positive Predictive Value; given that the problem was reported in the prior assessment(s), the proportion reporting that problem in the index assessment.

[h]NPV: Negative Predictive Value; given that the problem was not reported in the prior assessment(s), the proportion of index assessments that also did not report the problem.

Table 3. Area under the Receiver Operator Characteristic (ROC) curve for logistic models predicting each health indicator assuming different assessment frequencies.

	AUC[a]	95% CI
Moderate/Severe Depression [b]		
Weekly	0.9139	0.8931, 0.9348
Biweekly	0.8887	0.8655, 0.9119
Monthly	0.8873	0.8630, 0.9116
Baseline data only	0.7396	0.7010, 0.7782
Fair/Poor General Health		
Weekly	0.8840	0.8581, 0.9100
Biweekly	0.8758	0.8477, 0.9039
Monthly	0.8822	0.8543, 0.9101
Baseline data only	0.6760	0.6367, 0.7154
Poor Antidepressant Adherence		
Weekly	0.8396	0.8035, 0.8757
Biweekly	0.8268	0.7899, 0.8637
Monthly	0.8350	0.8000, 0.8701
Baseline data only	0.7578	0.7162, 0.7993
In Bed Due to Mental Health		
Weekly	0.7522	0.7058, 0.7986
Biweekly	0.6872	0.6358, 0.7385
Monthly	0.7197	0.6716, 0.7677
Baseline data only	0.6029	0.5542, 0.6515

[a]Area Under the Curve.

[b]PHQ-9 score ≥10.

Poor Antidepressant Adherence

While the overall predictive power was somewhat lower across models predicting reports of medication adherence problems, those models also showed that information collected biweekly or monthly was similar in its correlation with index assessment reports compared to information collected weekly (Table 3 and Figure 3). In Figure 3, the blue line represents weekly assessment attempts, the green line represents biweekly attempts, and the red line represents monthly attempts. The yellow line represents the ROC curve for the model predicting poor adherence using baseline data only. All other models also included baseline clinical and sociodemographic information.

The AUC for the model based on weekly assessments was 0.84 (95% CI 0.80, 0.88). The AUC for that model was not significantly different compared to either biweekly (P=.07) or monthly (P=.60) assessment attempts. In addition to prior information about patients' medication adherence, patients' age and baseline physical functioning consistently contributed to the predictive power of these models. Assuming weekly assessment attempts, the point on the ROC curve with the greatest number of assessments correctly classified (probability of adherence problems=.58) had a sensitivity of .86 and a specificity of .41.

Days in Bed

Models predicting days in bed due to mental health problem had the lowest predictive accuracy as measured by the AUC's for models based on weekly, biweekly, and monthly assessment attempts (Table 3 and Figure 4). In Figure 4, the blue line represents the ROC curve for the model based on weekly assessment attempts, the green line represents biweekly assessment attempts, and the red line represents monthly attempts. The yellow line represents the prediction with baseline data only, and all other models also included baseline clinical and sociodemographic information. While the AUC for weekly assessments was significantly different than either biweekly (P=.05) or monthly assessments (P=.05), there was no statistically significant difference in the AUC for biweekly and monthly calls, (P=.57). In addition to the patient's prior reports of days in bed, prior reports of depressive symptoms, as well as their baseline physical and mental functioning were significant predictors of days in bed.

Figure 1. Receiver Operator Characteristic (ROC) curves for models predicting patient reports of moderate/severe depression, as measured by a PHQ-9 score ≥10.
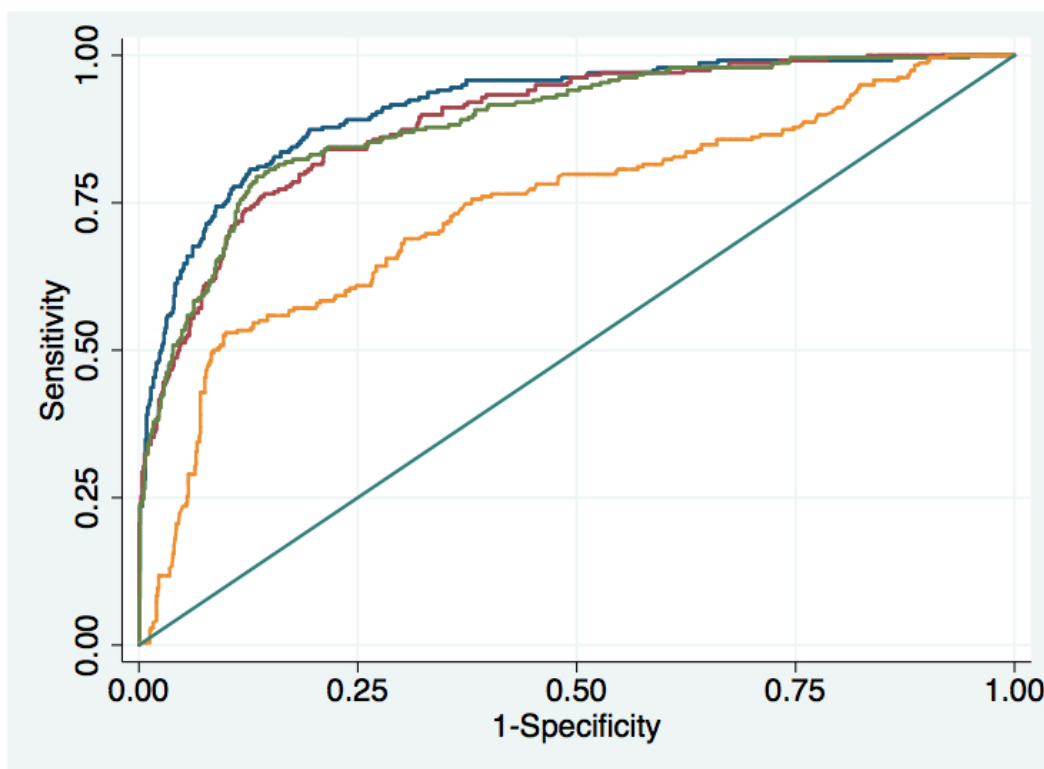
Figure 2. Receiver Operator Characteristic (ROC) curves for models predicting patient reports of fair or poor general health status.
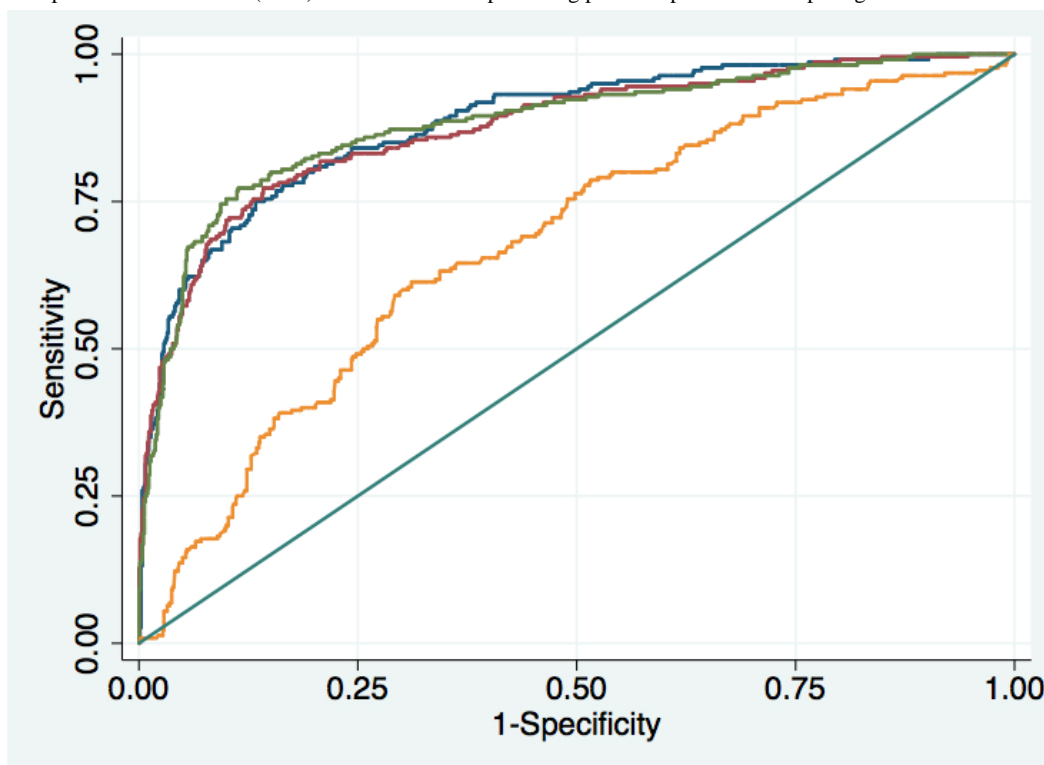
Figure 3. Receiver Operator Characteristic (ROC) curves for models predicting patient reports of poor antidepressant medication adherence.
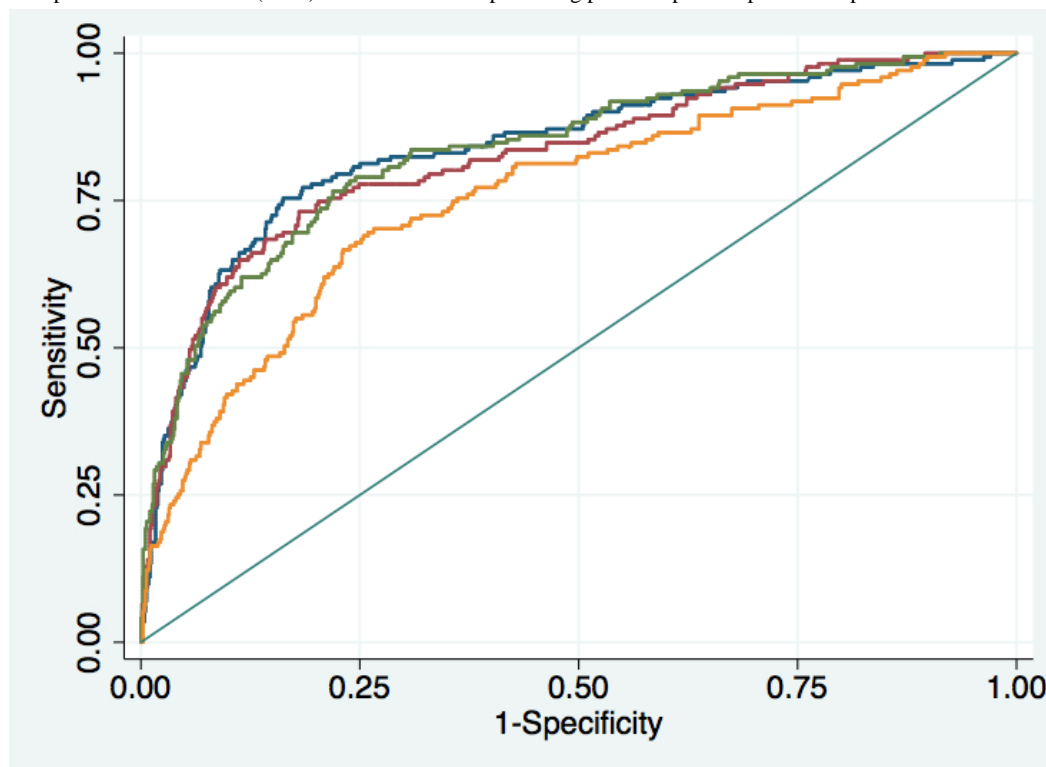
Figure 4. Receiver Operator Characteristic (ROC) curves for models predicting patient reports of being bedbound due to mental health problems.
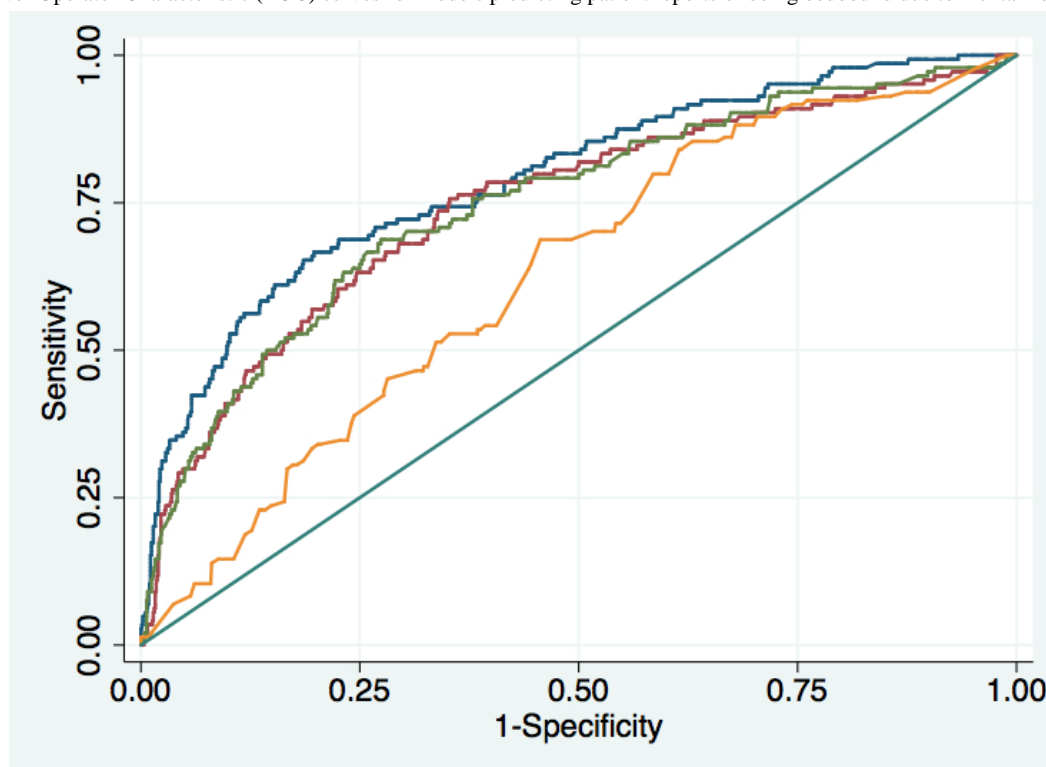
Discussion

Principal Findings

These analyses suggest that some IVR assessments of health and behavioral risk factors among patients with depression diagnoses may be unnecessary because patients' responses are predictable based on their prior pattern of reports. In particular,

we found that there is little to be gained from asking patients to report their PHQ-9 depression scores weekly and only a negligible incremental difference between biweekly and monthly assessment attempts. A similar pattern was observed with patients' reports of fair or poor perceived general health.

Less frequent assessments of a given health indicator, particularly when that indicator is measured via a multi-item

scale such as the PHQ-9, would have two benefits. First, it may be possible to decrease patients' response burden and risk for dropout by avoiding repetitive assessments of the same health problem. Second, by avoiding redundancy in IVR monitoring, more efficient messages could be designed that would cover a broader range of clinical parameters. In the current study, patients reported an average of more than two comorbid chronic conditions. Minimizing redundant questioning would allow for more comprehensive monitoring of comorbidities that may complicate the treatment of patients' depression and pose an independent threat to patients' health.

For two of the outcomes we examined—medication nonadherence and bed-bound status—prior IVR reports were only moderately successful in predicting patients' responses in a subsequent call. Several explanations are possible. It may be that adherence and days in bed were not reliably measured or that other still unmeasured predictors are more important in determining these health behaviors prospectively. Or it may be that these health indicators were in fact changing in unpredictable ways more rapidly than the frequency of monitoring could detect. If the latter reason is true, it may mean that even more frequent assessments are needed to detect all problems that arise. In any case, the approach to examining the frequency of monitoring presented here represents a framework for evaluating those options and making more informed choices about what health indicators to monitor and how often.

Assessments conducted in the current study were completed as part of a clinical service, with feedback to patients' primary care team and informal caregivers when serious problems were reported. It may be that those feedback reports led to interventions that stabilized patients' health status in ways that made subsequent patient reports more predictable. For obvious reasons, collecting patient health information without acting on it would be ethically challenging, but such information could provide insights into the appropriate periodicity of IVR monitoring for various outcomes. On the other hand, data used in the current study are more representative of what patients are likely to report in "real-world" practices, and the fact that we found that weekly assessments may produce redundant information is encouraging for health care organizations struggling with how best to manage their patients with multiple, competing health demands.

Patients who recently changed their antidepressant medication regimen may be more likely to experience side effects leading to adherence problems. The current system was not linked to pharmacy records. Such linkages represent an excellent example of the way in which monitoring systems that include a broader array of potential determinants of patients' health may help ensure that mobile health services focus on health indicators providing the most prognostically important information in the context of everything that is known about the patient.

Predictive models such as these could be used along with advanced machine learning algorithms to tailor the frequency of monitoring across patients, time, and health indicators. For example, time saved gathering redundant information about the trajectory of patients' depressive symptoms could be used to provide cognitive behavioral therapy designed to improve patients' mood by teaching skills such as cognitive restructuring or increased pleasurable activities [32]. Or for patients with depression and comorbid medical disorders, more efficient algorithms could adapt automatically in order to focus on the patient's other diseases, symptoms, or self-care behaviors that need greater attention to promote overall wellness. In brief, data mining approaches illustrated in the current study could be linked with algorithms that automatically update the content of patients' repeated mobile health interactions, maximizing the emphasis on patient education while continually monitoring the health problems that pose the greatest risk to patients' current and future risk for complications.

Each of the four outcomes examined could have been characterized using ordinal or even continuous measures, and the choice of dichotomizing the outcomes may have decreased the models' predictive power. We chose binary outcomes because clinical decisions (eg, whether to call the patient, request a visit, or change a prescription) are often binary, and these logistic models lend themselves to comparison via ROC curves that are familiar to many health care professionals. Nevertheless, data mining includes an increasingly large armamentarium of approaches that could be brought to bear on clinical prediction problems, depending on (for example) the functional form of the outcome, the amount of data available, and whether the relationship of interest is represented by "noisy" data generated from an underlying parametric model.

The current study used logistic regression, cross validation, and ROC curves to identify the predictive trends in patients' IVR-reported data. Artificial neural networks (ANNs) are an alternative parametric approach with more than 15 years of applications to medical diagnostics [33]. Support Vector Machines [34] represent a popular, nonparametric alternative to ANNs [35] for complex classification problems, particularly when the boundaries between groups (eg, between depressed and nondepressed patients) are irregular with respect to predictor variables and sufficient data are available for classification despite noise. Hierarchical latent-variable models (eg, Hidden-Markov Models [36]) could be used to capture underlying latent determinants of depression scores so that medical decisions can be conditioned on that latent information. If a continuous depression score were the outcome, moving average models with exponential smoothing could provide an initial understanding of data trends [37,38]. Other methods for modeling nonstationarities include autoregressive integrating moving averages (ARIMA) models [39] or regression-based forecasting models to extract complex characteristics of time series. More general models for state space representation also could be used to describe the motion of dynamic systems and extract position estimates as well as their derivatives eg, velocities or accelerations) from noisy data sources [40].

Regardless of the analytic approach, it may be that prediction of patients' responses could be improved by including more prior information in the prediction (eg, information from a larger number of prior IVR assessments). In the current study, we attempted to strike a balance between maximizing the predictive accuracy for a given patient, and including in the analyses a large, more representative sample of patients with a sufficient number of assessments (ie, by requiring no more than five prior

assessments with at least a 1-week, 2-week, and 1-month gap between each). Similar analyses in the context of data from large health plans may significantly improve the evidence base for clinical decision making.

Conclusions

In summary, the content and frequency of current mobile health assessments is almost entirely based on a fixed schedule and expert opinion, rather than being individualized based on patients' previously reported status. These analyses indicate that the technical feasibility of gathering high frequency health data may in some instances exceed the clinical benefit of doing so. In particular, weekly or biweekly depressive symptom reports may provide little marginal information regarding how the person is doing relative to collecting that information monthly. Data mining may allow us to detect trends in patient reports that can be used by intelligent systems to accurately predict patients' health status. The next generation of automated health assessment services should use these or other data mining techniques to avoid redundant assessments and gather data at the frequency that maximizes the value of the information collected. Such adaptive systems could be much more patient-friendly and could accommodate a much broader set of risk factors for the large and growing number of patients who have multiple chronic diseases.

Conflicts of Interest

None declared.

References

1. Kahn JG, Yang JS, Kahn JS. 'Mobile' health needs and opportunities in developing countries. Health Aff (Millwood) 2010 Feb;29(2):252-258 [FREE Full text] [doi: 10.1377/hlthaff.2009.0965] [Medline: 20348069]
2. Milne RG, Horne M, Torsney B. SMS reminders in the UK national health service: an evaluation of its impact on "no-shows" at hospital out-patient clinics. Health Care Manage Rev 2006;31(2):130-136. [Medline: 16648692]
3. Barclay E. Text messages could hasten tuberculosis drug compliance. Lancet 2009 Jan 3;373(9657):15-16. [Medline: 19125443]
4. Moore HK, Mundt JC, Modell JG, Rodrigues HE, DeBrota DJ, Jefferson JJ, et al. An examination of 26,168 Hamilton Depression Rating Scale scores administered via interactive voice response across 17 randomized clinical trials. J Clin Psychopharmacol 2006 Jun;26(3):321-324. [doi: 10.1097/01.jcp.0000219918.96434.4d] [Medline: 16702899]
5. González GM, Costello CR, La Tourette TR, Joyce LK, Valenzuela M. Bilingual telephone-assisted computerized speech-recognition assessment: is a voice-activated computer program a culturally and linguistically appropriate tool for screening depression in English and Spanish? Cult Divers Ment Health 1997;3(2):93-111. [Medline: 9231537]
6. Kobak KA, Taylor LH, Dottl SL, Greist JH, Jefferson JW, Burroughs D, et al. A computer-administered telephone interview to identify mental disorders. JAMA 1997 Sep 17;278(11):905-910. [Medline: 9302242]
7. Mundt JC, Kobak KA, Taylor LV, Mantle JM, Jefferson JW, Katzelnick DJ, et al. Administration of the Hamilton Depression Rating Scale using interactive voice response technology. MD Comput 1998;15(1):31-39. [Medline: 9458661]
8. Brodey BB, Rosen CS, Winters KC, Brodey IS, Sheetz BM, Steinfeld RR, et al. Conversion and validation of the Teen-Addiction Severity Index (T-ASI) for Internet and automated-telephone self-report administration. Psychol Addict Behav 2005 Mar;19(1):54-61. [doi: 10.1037/0893-164X.19.1.54] [Medline: 15783278]
9. Mundt JC, Moore HK, Bean P. An interactive voice response program to reduce drinking relapse: a feasibility study. J Subst Abuse Treat 2006 Jan;30(1):21-29. [doi: 10.1016/j.jsat.2005.08.010] [Medline: 16377449]
10. 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]
11. Bauer S, Moessner M. Technology-enhanced monitoring in psychotherapy and e-mental health. J Ment Health 2012 Aug;21(4):355-363. [doi: 10.3109/09638237.2012.667886] [Medline: 22548363]
12. Piette JD, Marinec N, Gallegos-Cabriales EC, Gutierrez-Valverde JM, Rodriguez-Saldaña J, Mendoz-Alevares M, et al. Spanish-speaking patients' engagement in interactive voice response (IVR) support calls for chronic disease self-management: data from three countries. J Telemed Telecare 2013 Mar 26 (forthcoming). [doi: 10.1177/1357633X13476234] [Medline: 23532005]

XSL•FO

RenderX

13.   Piette JD, Rosland AM, Marinec NS, Striplin D, Bernstein SJ, Silveira MJ. Engagement with automated patient monitoring and self-management support calls: experience with a thousand chronically ill patients. Med Care 2013 Mar;51(3):216-223. [doi: 10.1097/MLR.0b013e318277ebf8] [Medline: 23222527]

14.   Piette JD. Patient education via automated calls: a study of English and Spanish speakers with diabetes. Am J Prev Med 1999 Aug;17(2):138-141. [Medline: 10490057]

15.   Piette JD, McPhee SJ, Weinberger M, Mah CA, Kraemer FB. Use of automated telephone disease management calls in an ethnically diverse sample of low-income patients with diabetes. Diabetes Care 1999 Aug;22(8):1302-1309 [FREE Full text] [Medline: 10480775]

16.   Barnett K, Mercer SW, Norbury M, Watt G, Wyke S, Guthrie B. Epidemiology of multimorbidity and implications for health care, research, and medical education: a cross-sectional study. Lancet 2012 Jul 7;380(9836):37-43. [doi: 10.1016/S0140-6736(12)60240-2] [Medline: 22579043]

17.   Fortin M, Hudon C, Haggerty J, Akker M, Almirall J. Prevalence estimates of multimorbidity: a comparative study of two sources. BMC Health Serv Res 2010;10:111 [FREE Full text] [doi: 10.1186/1472-6963-10-111] [Medline: 20459621]

18.   Lichtenstein MJ, Sweetnam PM, Elwood PC. Visit frequency for controlled essential hypertension: general practitioners' opinions. J Fam Pract 1986 Oct;23(4):331-336. [Medline: 3760794]

19.   Schectman G, Barnas G, Laud P, Cantwell L, Horton M, Zarling EJ. Prolonging the return visit interval in primary care. Am J Med 2005 Apr;118(4):393-399. [doi: 10.1016/j.amjmed.2005.01.003] [Medline: 15808137]

20.   DeSalvo KB, Block JP, Muntner P, Merrill W. Predictors of variation in office visit interval assignment. Int J Qual Health Care 2003 Oct;15(5):399-405 [FREE Full text] [Medline: 14527983]

21.   DeSalvo KB, Bowdish BE, Alper AS, Grossman DM, Merrill WW. Physician practice variation in assignment of return interval. Arch Intern Med 2000 Jan 24;160(2):205-208. [Medline: 10647759]

22.   Cios KJ, Kacprzyk J. Medical Data Mining and Knowledge Discovery. New York, NY: Physica-Verlag; 2001.

23.   Cios KJ, Moore GW. Uniqueness of medical data mining. Artif Intell Med 2002;26(1-2):1-24. [Medline: 12234714]

24.   Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. J Gen Intern Med 2001 Sep;16(9):606-613 [FREE Full text] [Medline: 11556941]

25.   Löwe B, Unützer J, Callahan CM, Perkins AJ, Kroenke K. Monitoring depression treatment outcomes with the patient health questionnaire-9. Med Care 2004 Dec;42(12):1194-1201. [Medline: 15550799]

26.   Miilunpalo S, Vuori I, Oja P, Pasanen M, Urponen H. Self-rated health status as a health measure: the predictive value of self-reported health status on the use of physician services and on mortality in the working-age population. J Clin Epidemiol 1997 May;50(5):517-528. [Medline: 9180644]

27.   Grant MD, Piotrowski ZH, Chappell R. Self-reported health and survival in the Longitudinal Study of Aging, 1984-1986. J Clin Epidemiol 1995 Mar;48(3):375-387. [Medline: 7897459]

28.   DeSalvo KB, Bloser N, Reynolds K, He J, Muntner P. Mortality prediction with a single general self-rated health question. A meta-analysis. J Gen Intern Med 2006 Mar;21(3):267-275 [FREE Full text] [doi: 10.1111/j.1525-1497.2005.00291.x] [Medline: 16336622]

29.   Kroenke K, Strine TW, Spitzer RL, Williams JB, Berry JT, Mokdad AH. The PHQ-8 as a measure of current depression in the general population. J Affect Disord 2009 Apr;114(1-3):163-173. [doi: 10.1016/j.jad.2008.06.026] [Medline: 18752852]

30.   Ware J, Kosinski M, Keller SD. A 12-Item Short-Form Health Survey: construction of scales and preliminary tests of reliability and validity. Med Care 1996 Mar;34(3):220-233. [Medline: 8628042]

31.   Bradley AP. The use of the area under the ROC curve in the evaluation of machine learning algorithms. Pattern Recognition 1997 Jul;30(7):1145-1159. [doi: 10.1016/S0031-3203(96)00142-2]

32.   Naylor MR, Helzer JE, Naud S, Keefe FJ. Automated telephone as an adjunct for the treatment of chronic pain: a pilot study. J Pain 2002 Dec;3(6):429-438. [Medline: 14622728]

33.   Armoni A. Use of neural networks in medical diagnosis. MD Comput 1998;15(2):100-104. [Medline: 9540322]

34.   Cortes C, Vapnik V. Support-vector networks. Mach Learn 1995 Sep;20(3):273-297. [doi: 10.1007/BF00994018]

35.   Tonello L, Vescini F, Caudarella R. Support vector machines versus artificial neural network: Who is the winner? Kidney Int 2007 Jan;71(1):84-85. [doi: 10.1038/sj.ki.5001907] [Medline: 17167512]

36.   Fine S, Singer Y, Tishby N. The hierarchical hidden marcov model: analysis and applications. Machine Learning 1998;32:41-62.

37.   Spliid H. Monitoring medical procedures by exponential smoothing. Stat Med 2007 Jan 15;26(1):124-138. [doi: 10.1002/sim.2520] [Medline: 16479560]

38.   Verrier RL. Elevated T-wave alternans predicts nonsustained ventricular tachycardia in association with percutaneous coronary intervention in ST-segment elevation myocardial infarction (STEMI) patients. Journal of Cardiovascular Electophysiology 2013 (forthcoming).

39.   Sebestyen B, Rihmer Z, Balint L, Szokontor N, Gonda X, Gyarmati B, et al. Gender differences in antidepressant use-related seasonality change in suicide mortality in Hungary, 1998-2006. World J Biol Psychiatry 2010 Apr;11(3):579-585. [doi: 10.3109/15622970903397722] [Medline: 20218927]

40.   Lavieri MS, Puterman L, Tyldesley S, Morris WJ. When to treat prostate cancer patients based on their PSA dynamics. IIE Transactions on Healthcare Systems Engineering 2012;2:62-77.

Abbreviations

ANN: artificial neural networks
ARIMA: autoregressive integrating moving averages
AUC: area under the curve
IVR: interactive voice response calls
PHQ-9: 9-item Patient Health Questionnaire
ROC: Receiver Operator Characteristic curve

XSL•FO
RenderX

Original Paper

Assessing Adolescent Asthma Symptoms and Adherence Using Mobile Phones

Shelagh A Mulvaney[1,2,3], PhD; Yun-Xian Ho[2], PhD; Cather M Cala[4]; Qingxia Chen[2,5], PhD; Hui Nian[5], MS; Barron L Patterson[3], MD; Kevin B Johnson[2,3], MD, MS

[1]School of Nursing, Vanderbilt University School of Medicine, Nashville, TN, United States

[2]Department of Biomedical Informatics, Vanderbilt University School of Medicine, Nashville, TN, United States

[3]Department of Pediatrics, Vanderbilt University School of Medicine, Nashville, TN, United States

[4]School of Medicine, University of Alabama, Birmingham, AL, United States

[5]Department of Biostatistics, Vanderbilt University School of Medicine, Nashville, TN, United States

Corresponding Author:
Shelagh A Mulvaney, PhD
School of Nursing
Vanderbilt University School of Medicine
461 21st Ave
Nashville, TN, 37240
United States
Phone: 1 615 322 1198
Fax: 1 615 343 5898
Email: shelagh.mulvaney@vanderbilt.edu

Abstract

Background: Self-report is the most common method of measuring medication adherence but is influenced by recall error and response bias, and it typically does not provide insight into the causes of poor adherence. Ecological momentary assessment (EMA) of health behaviors using mobile phones offers a promising alternative to assessing adherence and collecting related data that can be clinically useful for adherence problem solving.

Objective: To determine the feasibility of using EMA via mobile phones to assess adolescent asthma medication adherence and identify contextual characteristics of adherence decision making.

Methods: We utilized a descriptive and correlational study design to explore a mobile method of symptom and adherence assessment using an interactive voice response system. Adolescents aged 12-18 years with a diagnosis of asthma and prescribed inhalers were recruited from an academic medical center. A survey including barriers to mobile phone use, the Illness Management Survey, and the Pediatric Asthma Quality of Life Questionnaire were administered at baseline. Quantitative and qualitative assessment of asthma symptoms and adherence were conducted with daily calls to mobile phones for 1 month. The Asthma Control Test (ACT) was administered at 2 study time points: baseline and 1 month after baseline.

Results: The sample consisted of 53 adolescents who were primarily African American (34/53, 64%) and female (31/53, 58%) with incomes US$40K/year or lower (29/53, 55%). The majority of adolescents (37/53, 70%) reported that they carried their phones with them everywhere, but only 47% (25/53) were able to use their mobile phone at school. Adolescents responded to an average of 20.1 (SD 8.1) of the 30 daily calls received (67%). Response frequency declined during the last week of the month (b=-0.29, P<.001) and was related to EMA-reported levels of rescue inhaler adherence (r= 0.33, P=.035). Using EMA, adolescents reported an average of 0.63 (SD 1.2) asthma symptoms per day and used a rescue inhaler an average of 70% of the time (SD 35%) when they experienced symptoms. About half (26/49, 53%) of the instances of nonadherence took place in the presence of friends. The EMA-measured adherence to rescue inhaler use correlated appropriately with asthma control as measured by the ACT (r=-0.33, P=.034).

Conclusions: Mobile phones provided a feasible method to assess asthma symptoms and adherence in adolescents. The EMA method was consistent with the ACT, a widely established measure of asthma control, and results provided valuable insights regarding the context of adherence decision making that could be used clinically for problem solving or as feedback to adolescents in a mobile or Web-based support system.

Introduction

Treatment for asthma is typically addressed initially through the use of what is known as a "rescue" inhaler that is used at the time that symptoms, such as shortness of breath, cough, or wheezing, occur. If symptoms persist over time despite use of rescue medications, a controller or "everyday" inhaler is prescribed. Thus, adherence to use of inhalers is a critical mediator of asthma control and health care utilization [1]. However, the primary method for measuring medication adherence, patient self-report, is hindered by recall error, response bias, and difficulty identifying specific contextual and psychosocial barriers to adherence [2]. In pediatric asthma, Jonasson and colleagues (1999) found that as adherence decreased (according to remaining doses in inhalers) traditional retrospective self-report became increasingly inaccurate [3]. Other research has documented overreporting of asthma control by adolescents using traditional retrospective self-report measures compared to clinician-rated estimates of asthma control [4].

A growing body of literature has documented progress in measurement of adherence and health behaviors through real-time or near real-time data capture using mobile technologies. This measurement method is known as ecological momentary assessment or EMA [2]. The method may reduce response bias introduced through social desirability and/or retrospective self-report and allows in vivo assessment of emotions, behaviors, and cognitions actually in, or near, the context in which they are experienced. This approach has been used to assess a wide variety of health behaviors and health-related conditions [5,6], pain [7,8], adolescent diabetes [9], and smoking [10], as well as to assess health information needs [11]. This method allows assessment of many aspects of health and health behaviors including physiological processes, timing or duration of events, patterns of symptoms over time, and situational and psychosocial conditions surrounding a health decision or event [2,12].

Psychosocial barriers to adherence for the adolescent age group often involve peers and feelings of stigma related to using an inhaler or taking medications [13]. Identification of the behavioral and situational correlates of adherence is necessary for clinicians and parents to engage the adolescent in focused problem solving. Mobile assessment has been qualitatively explored as a potential method to monitor asthma [14], but no research has explored this technology to assess adherence, symptoms, or used it to identify situational correlates of asthma adherence such as social context.

The primary goals of this research were to determine the feasibility of using EMA to assess asthma symptoms and adherence, identify contextual correlates of adherence using EMA, and compare EMA to the Asthma Control Test (ACT), a widely used self-report assessment of asthma control.

Methods

Participants

Participants were recruited from Vanderbilt University Medical Center. Adolescent candidates were included if they were between the ages of 12-18 years, had use of their own cell phone, were prescribed an inhaler, and had a diagnosis of asthma, as indicated by their parent or guardian (referred to as "parent").

Potential participants with a diagnosis of asthma were identified in the General Pediatrics clinic database by the director (co-author, BP) through use of ICD-9 (International Classification of Diseases, ninth revision) codes and reviews of patient medication lists. All participants were receiving primary care through the Vanderbilt Medical Center. We used three recruitment strategies. First, the research team sent letters to parents of adolescents (N=515) determined to be eligible for participation based on a diagnosis of asthma present in the adolescents' medical records. If an adolescent was 18 years or older, the letter was addressed directly to him/her. Following the letters, parents of the adolescent candidate were contacted to determine interest. Second, we distributed flyers and interest cards to the waiting areas of General Pediatrics, Adolescent Medicine, and Pediatric Pulmonology clinics, as well as the pediatric emergency department. A research assistant contacted candidates who completed and submitted an interest card. Finally, we posted an announcement to a research notification listserv and in a medical center online newsletter. A total of 199 candidate parent-adolescent dyads were assessed for eligibility via telephone call with the parent. We identified 83 candidates out of 199 (42%) who were not eligible to participate or not interested in participating and 116 candidates (58%) who were eligible to participate. Candidates were primarily excluded because the adolescent did not have their own cell phone (22/83, 27%) or lacked interest in the study (28/83, 34%). Out of those who were eligible, 54 (47%) were enrolled and 62 (53%) did not progress to baseline because a parent could not be reached, or the family did not show up for their initial appointment. Out of the candidates who were enrolled, 1 participant did not complete the study. The final dataset consisted of 53 adolescent-guardian dyads that completed the study.

Procedures and Measures

At the initial session, adolescents were informed that parents and clinicians would not have access to their study data, nor would any indication of their participation be included in their medical record. At baseline, adolescents completed questions generated for this research related to mobile phone use and possible barriers to adolescent mobile phone use for asthma (ie, where they were able to use their phones, typical response to calls, need to earn their phone use), the ACT, the Illness Management Survey (IMS), and the Pediatric Asthma Quality of Life Questionnaire (PAQLQ). The ACT is a widely used

5-item measure of symptoms and rescue inhaler use over the previous 4 weeks [15]. The measure has been validated for use over the telephone [16]. We also included an item to assess the use of a daily controller inhaler over the previous 4 weeks. Parents completed the ACT as it related to their adolescent at baseline. Higher scores indicate better asthma control.

The IMS consists of 27 items that measure perceptions of barriers to adherence in adolescents with chronic illness including interactions with health care providers, cognitive abilities, family/peer influences, and denial of the illness. Responses range from strongly disagree (1) to strongly agree (5). The measure has been validated in adolescents with asthma and has demonstrated adequate internal reliability (Cronbach alpha=.84) [17]. Higher scores indicate greater perceived barriers to adherence.

The PAQLQ is a widely used measure of the perception of the impact of asthma on daily life. The PAQLQ is related to an objective measure of asthma control (spirometry) and has adequate psychometric properties cross-sectionally and at test-retest [18]. We used the brief version of the instrument with 13 items. Higher scores indicate greater perceived quality of life. All survey data were collected and managed using the online survey system Research Electronic Data Capture (REDCap) [19]. Parent consent and adolescent assent were obtained before study procedures commenced. The Institutional Review Board approved all study procedures.

Following baseline procedures, EMA was used to measure symptoms and adherence. An interactive voice response system (Telesage) was administered through the adolescents' mobile phones. At baseline, each adolescent completed a practice call on their cell phone with the researcher present to ensure they understood how to respond to questions. Each participant received 1 call per day for 30 days. The calls were scheduled for the family's preferred time between 6-8 p.m. in the evening. Adolescents could return a call to the system ("incoming") before midnight that day if they missed the "outgoing" call to them. The automated adaptive phone survey included between 2-9 questions depending on responses to the questions and lasted between 15-90 seconds. Figure 1 shows the adaptive EMA items. Adolescents responded to numeric or yes/no items using the keypad on their mobile phone and qualitative questions by voice.

After 1 month of EMA, adolescents and their guardians were asked to complete the ACT using REDCap at this second time point ("Time 2").

Statistical Analyses

To calculate rescue inhaler adherence for each participant using EMA, we divided the number of days the inhaler was used by the number of days asthma symptoms were reported. Adolescents who reported having no asthma symptoms did not receive a rescue adherence score. To calculate controller (everyday) adherence for the EMA method, we divided the number of days participants reported using their controller inhaler by the total number of daily call responses. A paired t test was used to determine if there were any differences between adolescents' ACT scores reported at baseline and Time 2 and their parents' reported scores. Data from the ACT administered at Time 2 were used for these comparisons in order to correspond to the calling period. Only EMA-measured rescue inhaler use was compared to the ACT, as it measures control of asthma related to rescue inhaler use. Within-subject comparisons of adherence and Spearman correlations were conducted to examine the relationship between EMA adherence and other self-report measures. Statistics were carried out using Stata v10 and R v2.13.1.

Figure 1. Ecological momentary assessment administered through mobile phones.
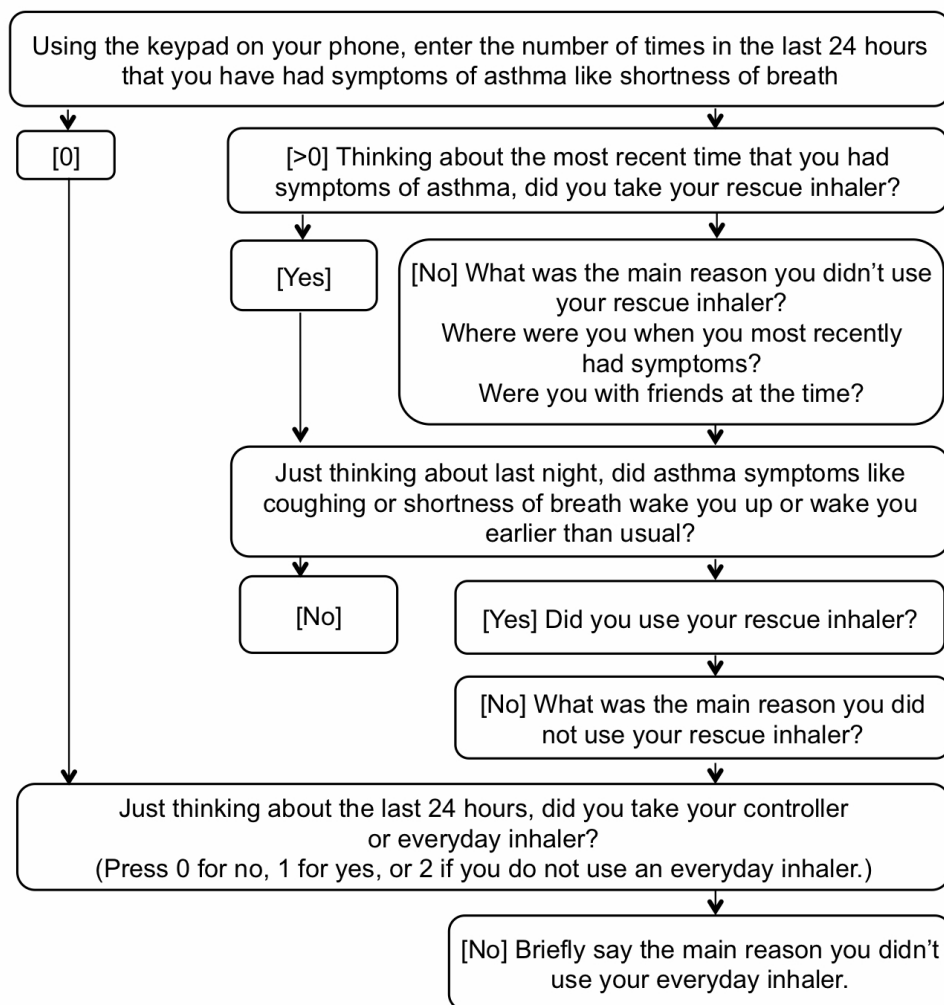
Results

Characteristics of the sample are in Table 1. The sample was primarily female (31/53, 58%), African American (33/53, 64%), and mid- to low-income with an annual household income at or below US$40,000 (29/53, 55%). The mean IMS score for assessing barriers to adherence was 2.7 out of 5 (SD 0.6) with a lower score indicating lower barriers. Mean PAQLQ score was 5.5 (SD 1.4) out of a maximum of 7. Mean PAQLQ Symptoms Scale score was 5.4 (SD 1.3). The mean total ACT score at baseline was 19.8 (SD 4.0, range 9-25), and at Time 2 was 19.4 (SD 4.1, range 11-25). There were no statistically significant differences between adolescent baseline and Time 2 scores (P=.36) or between adolescent and parent ACT scores at either baseline or Time 2 (P=.36 and .70, respectively).

Cell Phone Use

Adolescents were very confident about how to use their cell phones (mean 9.1, SD 1.8, range 1-10). Just over half (31/53, 58%) were able to access the Internet with their phones. The majority (50/53, 94%) reported that their school had restrictions on using mobile phones at school with 38% (18/47) not allowed to use the phone at all, and 47% (22/47) of those allowed to use the phone being able to do so only outside of class. The majority of adolescents (28/53, 53%) were not sure if their school would allow mobile phone use for asthma, 23% (12/53) were not allowed, and 25% (13/53) could use their phones specifically for asthma. Most felt that school mobile phone policies did not impact how they took care of asthma (46/53, 87%). Outside of school, 70% of adolescents (37/53) reported that they always take their cell phones with them everywhere they go, and 57% (30/53) of adolescents reported that they immediately try to see who contacted them upon receiving a call or message. About one-third of adolescents were required to earn or pay for their cell phone plans (16/53, 30%). Adolescents reported relatively little use of their phones for asthma, such as accessing online information (9/53, 17%), sending a text message (9/53, 17%), setting a reminder to do something about asthma (17/53, 32%), calling a family member (21/53, 40%), or talking to a nurse or

doctor (10/53, 19%). None reported using a smartphone application for asthma.

## Disposition of Daily Calls and Missing Data

Figure 2 summarizes the disposition of daily calls. Figure 2 shows that there were 69.9% (1063/1520) of calls with data available for analyses. This included outgoing and incoming calls (returned outgoing calls). About 6% (67/1130) of the calls were excluded due to dropped calls or hang-ups.

Figure 3 shows that an average of 70% of calls each day over the study period contained complete data. Adolescents had an average of 20.1 (SD 8.1) responses or 67% (SD 27.7) of all daily call responses each. There was some decline in calls with data over time (b=-0.29, P<.001). Missingness was defined as the number of days of no response to the calls. Missingness was not correlated with age, gender, household income, asthma control as measured by the ACT or controller adherence as measured by EMA (P=.86, .41, .10, .94, .12, respectively). However, missingness was significantly related to EMA-measured rescue inhaler adherence (r=-.33, P=.04), indicating that as the number of missed calls increased, adherence decreased.

Table 1. Characteristics of the sample (n=53).

Characteristics		Mean (SD) or n (%)
Age		15.2 (1.7)
Gender		
	Male	22 (42%)
	Female	31 (58%)
Ethnicity		
	White/Caucasian	18 (34%)
	African American	33 (62%)
	Hispanic	2 (4%)
Type of school		
	Public	49 (92%)
	Private	2 (4%)
	Home-schooled	2 (4%)
Home life		
	Single parent	26 (49%)
	Both parents	23 (43%)
	Other legal guardian	4 (8%)
Household income (USD)		
	Less than $20,000	17 (32%)
	$20,001-$40,000	12 (23%)
	$40,001-$70,000	12 (23%)
	More than $70,000	8 (15%)
	Decline to answer	4 (8%)
Guardian's education level		
	Grade 7-11	8 (15%)
	Grade 12	11 (21%)
	Some college, no degree	16 (30%)
	College degree	13 (25%)
	Graduate school	5 (9%)
Asthma medication regimen		
	Rescue inhaler	53 (100%)
	Controller & rescue inhaler	23 (43%)

Figure 2. Flow diagram of daily EMA calls.
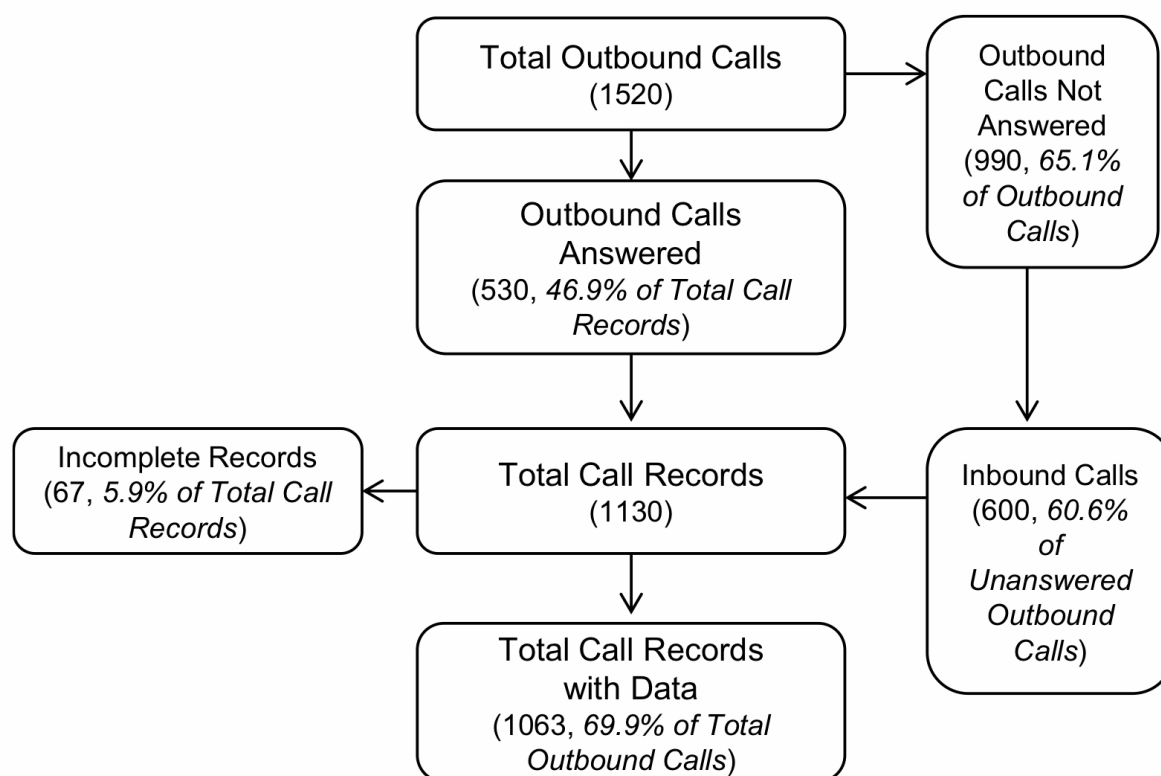
Symptoms, Adherence, and Contextual Characteristics

Of the 1063 calls with data, 27% (287/1063) of those calls recorded symptoms of asthma during the previous 24 hours. During the study period, 12 of the 53 (23%) adolescents reported having no asthma symptoms. When adolescents reported a symptom, they reported on average using a rescue inhaler 70% (SD 35) of the time. The most common reasons reported by adolescents for not using the rescue inhaler included not needing the inhaler (23/57, 40%) or not having it (17/57, 30%). When symptoms were experienced, adolescents were most commonly at home (18/51, 32%), at school (8/51, 16%), or exercising (6/51, 12%). Adolescents reported being in the company of friends 53% (29/55) of the time that the rescue inhaler was not used in the context of symptoms. Controller inhaler use averaged 80% of the time over the study period (SD 29). The most common reason adolescents reported for not using the controller was they felt they did not need it (135/207, 65%).

Relationship Between EMA and ACT

Table 2 shows summary statistics for and bivariate correlations between the ACT and EMA. The ACT total and EMA means were appropriately negatively correlated indicating that as symptoms and use of a rescue inhaler increased, asthma control decreased. Individual ACT and EMA items also exhibited significant positive correlations on the three corresponding items related to shortness of breath, nighttime symptoms, and rescue inhaler use.

Figure 3. Percent of calls with data per day over 30 days (Dark gray=outbound calls; light gray=inbound calls).
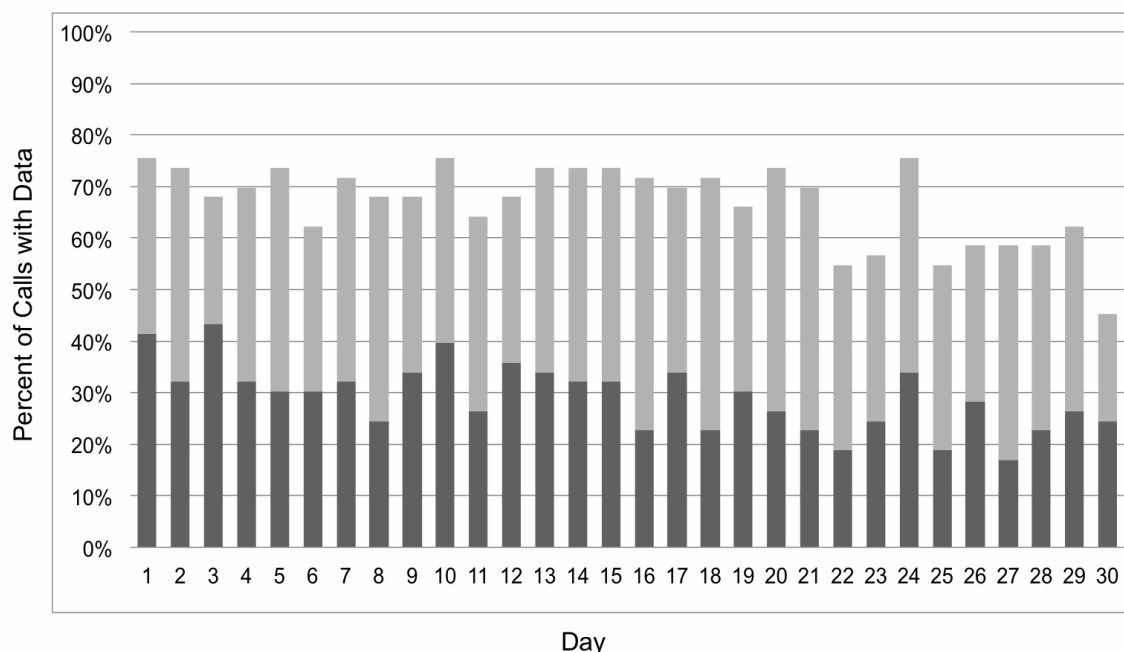
Table 2. Comparisons of EMA-measured rescue inhaler adherence and ACT scores for totals and corresponding items.

EMA	Median, mean (SD)	ACT	Median, mean (SD)	r [a]
Rescue inhaler adherence	0.83, 0.70 (0.35)	Total score for all 5 items[b]	19, 18.5 (4.0)	-0.33[c]
Using the keypad on your phone, enter the number of times, in the last 24 hours, that you have had symptoms of asthma, like shortness of breath or coughing.	0.25, 0.63 (1.2)	During the past 4 weeks, how often have you had shortness of breath?	2, 2.5 (1.3)	0.44[d]
Just thinking about last night, did asthma symptoms like coughing or shortness of breath wake you up, or wake you earlier than usual? [c]	0, 0.06 (0.10)	During the past 4 weeks, how often did your asthma symptoms (wheezing, coughing, shortness of breath, tightness or pain) wake you up at night or earlier than usual in the morning?	1, 1.8 (1.2)	0.65[d]
Thinking about the most recent time that you had symptoms of asthma in the last 24 hours, did you take your rescue inhaler?[e]	0.12, 0.20 (0.20)	During the past 4 weeks, how often have you used your rescue inhaler or nebulizer medication (such as albuterol)?	2, 2.4 (1.2)	0.52[d]

[a]EMA responses scaled to ACT for comparison.

[b]Total ACT score ranges from 5 to 25.

[c]P=.034.

[d]P<.001.

[e]"Yes" scored as 1, "No" scored as 0.

Discussion

Principal Results

We investigated the feasibility of a mobile method of measuring asthma symptoms and adherence in adolescents over time and compared that to a widely used self-report measure of asthma control. Regarding access and use of mobile phones, we determined that neither mobile phone accessibility nor need for immediate response were barriers to utilizing the phone for health assessment. Just over one third of adolescents were not allowed to use their phones at school at all, but the majority did not view school policies as a barrier to taking care of asthma. Interestingly, only a minority of the sample reported currently using their mobile phone to help with asthma (setting reminders, text messages, use of websites, or mobile apps), and 42% did not have Internet access through their phone. Because school access was unclear, we scheduled daily calls for the evening hours. It appears that real-time assessment and support for asthma using mobile phones could be hindered by current public school policies. In order to address this barrier, it may be necessary for researchers, parents, and advocacy groups to collaborate with school administrators to facilitate the use of mobile health devices for students.

Missing data and response burden are potential issues in daily or momentary health behavior assessments. On average, there was adequate response to the daily calls. Allowing incoming calls greatly enhanced the daily response levels. There was a decline in response to the calls over one month, with the 4th week the point at which that decline was observable. Although a 30-day period was selected to correspond to the established asthma control self-report measure, a 10-20 day assessment period may be optimal for once-a-day assessments in this population [20]. Similar to a previous study in asthma that assessed peak flow values [5], we found a wide range of response levels. It is currently unclear what number of observations should be used to provide the best sampling of asthma symptoms and behaviors. We anticipate that a threshold for adequate sampling will vary by goal of the research and exposure to asthma triggers. Interestingly, momentary assessment using self-report was still susceptible to response bias in terms of missing data. Higher levels of missing data were related to lower adherence. This speaks to the need for unobtrusive or passive methods of adherence assessment.

Measurement of symptoms and adherence using the EMA method provided specific data that may be useful for patient and clinical pattern recognition, problem solving, and decision making at both the population and individual patient levels. For example, the average number of symptoms per day was 0.6 but individual numbers ranged greatly. This method could be used as part of a baseline or periodic monitoring of symptoms during seasonal exposure to asthma triggers or after modification of a regimen. Additionally, insights regarding the reasons for inadequate adherence are difficult to assess clinically. The location-based data obtained for symptom occurrence were not particularly revealing indicating home, school, and sports events as relevant. However, the presence of peers was a prevalent and potentially impactful barrier to adherence in about half of the instances when a rescue inhaler was not used. Of particular interest was that the most common reason adolescents reported not using their rescue or controller inhalers was because they did not believe they needed it, despite the presence of asthma symptoms. These qualitative findings strongly suggest a valuable focus for patient education and adherence problem solving.

Limitations

Generalizability of our results is somewhat limited by the small sample of this pilot study. The daily calls could have positively influenced adherence through enhanced self-monitoring and greater awareness of symptoms and adherence. However, previous research in adolescents with diabetes did not indicate an EMA monitoring effect [9]. Similarly, a monitoring effect does not appear to have occurred in this study as the baseline and 1-month ACT scores were equivalent. Additional research will be needed to directly compare mobile methods of assessing asthma symptoms and adherence to other assessment methods, such as parent/spouse report, days of school missed, and other objective methods not directly related to the process of mobile data capture. Finally, we included a modified version of the ACT to allow parallel items for parent report of adolescent asthma control. That measure has not been validated.

Conclusions

The mobile measurement method described here provided a feasible means to probe patterns of symptoms and adherence over time and provided additional qualitative insights regarding contextual reasons for adherence problems. The use of this and related methods may be more expensive to implement compared to the traditional retrospective self-report [21] and is still susceptible to bias in missing data, but should be explored for its added value in clinical practice and integrated with tailored mobile intervention techniques to improve adherence.

Conflicts of Interest

None declared.

References

1. Rohan J, Drotar D, McNally K, Schluchter M, Riekert K, Vavrek P, et al. Adherence to pediatric asthma treatment in economically disadvantaged African-American children and adolescents: an application of growth curve analysis. J Pediatr Psychol 2010 May;35(4):394-404 [FREE Full text] [doi: 10.1093/jpepsy/jsp074] [Medline: 19710251]
2. Shiffman S, Stone AA, Hufford MR. Ecological momentary assessment. Annu Rev Clin Psychol 2008;4:1-32. [Medline: 18509902]
3. Jónasson G, Carlsen KH, Sødal A, Jonasson C, Mowinckel P. Patient compliance in a clinical trial with inhaled budesonide in children with mild asthma. Eur Respir J 1999 Jul;14(1):150-154 [FREE Full text] [Medline: 10489843]
4. Britto MT, Byczkowski TL, Hesse EA, Munafo JK, Vockell AL, Yi MS. Overestimation of impairment-related asthma control by adolescents. J Pediatr 2011 Jun;158(6):1028-1030.e1. [doi: 10.1016/j.jpeds.2011.01.034] [Medline: 21392791]
5. Ryan D, Cobern W, Wheeler J, Price D, Tarassenko L. Mobile phone technology in the management of asthma. J Telemed Telecare 2005;11 Suppl 1:43-46. [doi: 10.1258/1357633054461714] [Medline: 16035991]

6. Anhøj J, Møldrup C. Feasibility of collecting diary data from asthma patients through mobile phones and SMS (short message service): response rate analysis and focus group evaluation from a pilot study. J Med Internet Res 2004 Dec 2;6(4):e42 [FREE Full text] [doi: 10.2196/jmir.6.4.e42] [Medline: 15631966]

7. Jamison RN, Raymond SA, Slawsby EA, McHugo GJ, Baird JC. Pain assessment in patients with low back pain: comparison of weekly recall and momentary electronic data. J Pain 2006 Mar;7(3):192-199. [doi: 10.1016/j.jpain.2005.10.006] [Medline: 16516825]

8. Levin E, Levin A. Evaluation of spoken dialogue technology for real-time health data collection. J Med Internet Res 2006;8(4):e30. [doi: 10.2196/jmir.8.4.e30]

9. Mulvaney SA, Rothman RL, Dietrich MS, Wallston KA, Grove E, Elasy TA, et al. Using mobile phones to measure adolescent diabetes adherence. Health Psychol 2012 Jan;31(1):43-50. [doi: 10.1037/a0025543] [Medline: 21967662]

10. Mermelstein R, Hedeker D, Flay B, Shiffman S. Real-Time Data Capture and Adolescent Cigarette Smoking: Moods and Smoking. In: Stone AA, Shiffman S, Atienza AA, Nebeling L, editors. The Science of Real-Time Data Capture: Self-Reports in Health Research. New York: Oxford University Press; 2007:117-135.

11. Schnall R, Okoniewski A, Tiase V, Low A, Rodriguez M, Kaplan S. Using text messaging to assess adolescents' health information needs: an ecological momentary assessment. J Med Internet Res 2013;15(3):e54 [FREE Full text] [doi: 10.2196/jmir.2395] [Medline: 23467200]

12. Heron KE, Smyth JM. Ecological momentary interventions: incorporating mobile technology into psychosocial and health behaviour treatments. Br J Health Psychol 2010 Feb;15(Pt 1):1-39 [FREE Full text] [doi: 10.1348/135910709X466063] [Medline: 19646331]

13. Riekert K, Drotar D. Adherence to medical treatment in pediatric chronic illness: critical issuesanswered questions. In: Drotar D, editor. Promoting adherence to medical treatment in chronic childhood illness. Mahwah, NJ: Lawrence Erlbaum; 2000.

14. Pinnock H, Slack R, Pagliari C, Price D, Sheikh A. Understanding the potential role of mobile phone-based monitoring on asthma self-management: qualitative study. Clin Exp Allergy 2007 May;37(5):794-802. [doi: 10.1111/j.1365-2222.2007.02708.x] [Medline: 17456228]

15. Nathan RA, Sorkness CA, Kosinski M, Schatz M, Li JT, Marcus P, et al. Development of the asthma control test: a survey for assessing asthma control. J Allergy Clin Immunol 2004 Jan;113(1):59-65. [doi: 10.1016/j.jaci.2003.09.008] [Medline: 14713908]

16. Schatz M, Zeiger RS, Drane A, Harden K, Cibildak A, Oosterman JE, et al. Reliability and predictive validity of the Asthma Control Test administered by telephone calls using speech recognition technology. J Allergy Clin Immunol 2007 Feb;119(2):336-343. [doi: 10.1016/j.jaci.2006.08.042] [Medline: 17194469]

17. Logan D, Zelikovsky N, Labay L, Spergel J. The Illness Management Survey: identifying adolescents' perceptions of barriers to adherence. J Pediatr Psychol 2003 Sep;28(6):383-392 [FREE Full text] [Medline: 12904450]

18. Juniper EF, Guyatt GH, Feeny DH, Ferrie PJ, Griffith LE, Townsend M. Measuring quality of life in children with asthma. Qual Life Res 1996 Feb;5(1):35-46. [Medline: 8901365]

19. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap)--a metadata-driven methodology and workflow process for providing translational research informatics support. J Biomed Inform 2009 Apr;42(2):377-381 [FREE Full text] [doi: 10.1016/j.jbi.2008.08.010] [Medline: 18929686]

20. Jentzsch NS, Camargos PA, Colosimo EA, Bousquet J. Monitoring adherence to beclomethasone in asthmatic children and adolescents through four methods. Allergy 2009 Oct;64(10):1458-1462. [doi: 10.1111/j.1398-9995.2009.02037.x] [Medline: 19416142]

21. Ryan D, Price D, Musgrave SD, Malhotra S, Lee AJ, Ayansina D, et al. Clinical and cost effectiveness of mobile phone supported self monitoring of asthma: multicentre randomised controlled trial. BMJ 2012 Mar 23;344(1):e1756. [doi: 10.1136/bmj.e1756]

Abbreviations

ACT: Asthma Control Test
AHRQ: Agency for Healthcare Research and Quality
EMA: ecological momentary assessment
ICD-9: International Classification of Diseases, 9th revision
IMS: Illness Management Survey
PAQLQ: Pediatric Asthma Quality of Life Questionnaire
REDCap: Research Electronic Data Capture

XSL•FO
RenderX

XSL•FO
RenderX

Review

Scoping Review on Search Queries and Social Media for Disease Surveillance: A Chronology of Innovation

Theresa Marie Bernardo[1], DVM, MSc; Andrijana Rajic[2,3,4], DVM, MSc, PhD; Ian Young[2,3], BASc, PhD; Katie Robiadek[5], BA, MA; Mai T Pham[2,3], BSc, BASc, MSc; Julie A Funk[1], DVM, MSc, PhD

[1]College of Veterinary Medicine, Michigan State University, East Lansing, MI, United States

[2]Department of Population Medicine, University of Guelph, Guelph, ON, Canada

[3]Laboratory for Foodborne Zoonoses, Public Health Agency of Canada, Guelph, ON, Canada

[4]Agriculture Department, Food and Agriculture Organisation, Rome, Italy

[5]Department of Political Science, University of Wisconsin-Madison, Madison, WI, United States

Corresponding Author:
Theresa Marie Bernardo, DVM, MSc
College of Veterinary Medicine
Michigan State University
Veterinary Medical Center
784 Wilson Road Room G338
East Lansing, MI, 48824-1314
United States
Phone: 1 517 432 1292
Fax: 1 517 432 2937
Email: TheresaBernardo@gmail.com

Abstract

Background: The threat of a global pandemic posed by outbreaks of influenza H5N1 (1997) and Severe Acute Respiratory Syndrome (SARS, 2002), both diseases of zoonotic origin, provoked interest in improving early warning systems and reinforced the need for combining data from different sources. It led to the use of search query data from search engines such as Google and Yahoo! as an indicator of when and where influenza was occurring. This methodology has subsequently been extended to other diseases and has led to experimentation with new types of social media for disease surveillance.

Objective: The objective of this scoping review was to formally assess the current state of knowledge regarding the use of search queries and social media for disease surveillance in order to inform future work on early detection and more effective mitigation of the effects of foodborne illness.

Methods: Structured scoping review methods were used to identify, characterize, and evaluate all published primary research, expert review, and commentary articles regarding the use of social media in surveillance of infectious diseases from 2002-2011.

Results: Thirty-two primary research articles and 19 reviews and case studies were identified as relevant. Most relevant citations were peer-reviewed journal articles (29/32, 91%) published in 2010-11 (28/32, 88%) and reported use of a Google program for surveillance of influenza. Only four primary research articles investigated social media in the context of foodborne disease or gastroenteritis. Most authors (21/32 articles, 66%) reported that social media-based surveillance had comparable performance when compared to an existing surveillance program. The most commonly reported strengths of social media surveillance programs included their effectiveness (21/32, 66%) and rapid detection of disease (21/32, 66%). The most commonly reported weaknesses were the potential for false positive (16/32, 50%) and false negative (11/32, 34%) results. Most authors (24/32, 75%) recommended that social media programs should primarily be used to support existing surveillance programs.

Conclusions: The use of search queries and social media for disease surveillance are relatively recent phenomena (first reported in 2006). Both the tools themselves and the methodologies for exploiting them are evolving over time. While their accuracy, speed, and cost compare favorably with existing surveillance systems, the primary challenge is to refine the data signal by reducing surrounding noise. Further developments in digital disease surveillance have the potential to improve sensitivity and specificity, passively through advances in machine learning and actively through engagement of users. Adoption, even as supporting systems for existing surveillance, will entail a high level of familiarity with the tools and collaboration across jurisdictions.

XSL•FO
RenderX

Introduction

Social media and search behavior produce vast new data sources of largely untapped scientific potential. The threat of a global pandemic posed by outbreaks of influenza H5N1 (1997) and Severe Acute Respiratory Syndrome (SARS, 2002), both diseases of zoonotic origin, provoked interest in improving early warning systems and reinforced the need for combining data from different sources. It led to novel ideas, for example, the use of search query data from search engines such as Google [1,2] and Yahoo! [3] as an indicator of when and where influenza was occurring. This methodology has subsequently been extended to other diseases and has led to experimentation with new types of social media for disease surveillance as they have become available. Despite the emergence of disease surveillance as an innovative use of social media and search engine technologies, there is limited knowledge regarding the scope and efficacy of this novel application. With the potential to greatly improve disease surveillance and mitigation, there is a significant need to understand key chronological developments of the tools and methodologies in order to inform future endeavors and to assess this technology application for potential end-users.

Traditional narrative literature reviews provide useful overviews of broad research fields; however, their utility to inform policy and decision making is limited due to the lack of methodological transparency in terms of study selection and possible bias in interpretation [4,5]. Scoping reviews are a structured and formal knowledge synthesis method that can be used to rapidly identify, characterize, and contextualize existing knowledge and gaps in research [6-8]. They represent a relatively new methodology that has increasingly been adopted in health and various other sectors [6], including recent applications in food safety and zoonotic public health [8-10]. The objective of this scoping review was to formally assess the current state of knowledge regarding the use of online search queries and social media for disease surveillance in order to inform and encourage future work on early detection and more effective mitigation of the effects of foodborne illness. We used structured scoping review methods to identify, characterize, and evaluate all published primary research, expert review, and commentary articles investigating or discussing the use of social media in surveillance of infectious diseases. The results are presented and discussed within the context of existing research knowledge, as well as the surveillance and policy needs, gaps, and opportunities.

Methods

Review Protocol and Team Expertise

The review was informed by an ongoing scoping review protocol that includes details of the review methodology, definitions, and all forms used in the project (see Multimedia Appendix 1).

The review team consisted of all 6 co-authors with multidisciplinary expertise in epidemiology, infectious diseases, food safety and zoonoses, social media, and knowledge synthesis methods. An advisory committee consisting of 23 professionals from 12 government, academic, and civil society organizations and with expertise in epidemiology, food safety, risk communication, social media, spatial geography, computer science, and mathematics, was consulted throughout the review to ensure that relevant articles in their respective fields had not been missed. Preliminary results of the scoping review were presented to the advisory committee and stakeholder feedback was received at a related project initiative [11].

Review Question and Scope

The review question was "What is the current state of knowledge about the use and efficacy of mining social media text and Web query trends for disease surveillance?" Social media were defined as a group of Internet-based online and mobile applications (eg, Twitter, Facebook) that allow the creation and exchange of user-generated content and data [12]. Disease surveillance was defined as the ongoing systematic collection and analysis of data and the provision of information that leads to action being taken to prevent and control a disease [13]. This included activities related to early detection, prevention, control, and eradication of sporadic cases and outbreaks, endemic and epidemic diseases, and infectious and chronic diseases. Threats were limited to biological (viruses, parasites, bacteria, and their toxins) and chemical agents (melamine, pesticides).

Search Strategy

A pretested electronic search strategy was implemented in SciVerse Scopus (2002-2011) on August 16, 2011 (see Multimedia Appendix 1). The search strategy used a targeted combination of 17 social media and Internet-based tool terms (eg, blog, Internet), five disease terms (eg, outbreak), and five surveillance terms (eg, monitor). The search was limited to 2002 and onward to coincide with the wide use of Web 2.0 applications. A Scopus and Google Web search were also conducted to identify grey literature (eg, reports and newspaper articles); both were limited to the 100 most relevant hits. The Scopus Web search used the same search strategy as above, while the Google search used the query "social media for disease surveillance". The reference lists of 11 topic-related articles were hand-searched to identify any additional relevant citations potentially missed by the initial search strategy.

Scoping Review Management and Form Pretesting

All references were imported into the online bibliographic management program RefWorks and subsequently imported into DistillerSR, a Web-based systematic review software for relevance screening and data characterization and extraction.

Relevance screening and data characterization and extraction forms were pretested and refined to standardize interpretation among 4 reviewers before use. The relevance screening form

was pretested on 20 abstracts by 5 reviewers (TB, AR, KR, MP, and JF), and reviewing proceeded when kappa agreements were >0.7. The data characterization and extraction form was also pretested on five articles by 3 reviewers (TB, KR, and JF). A high agreement and only minor editorial discrepancies were observed for a couple of open-ended questions. These were discussed among the team members and the most practical yet robust data characterization and extraction process was determined.

Relevance Screening and Inclusion Criteria

Each abstract was screened for relevance against the inclusion and exclusion criteria by 2 independent reviewers (AR and JF, KR and TB). Any peer or non–peer-reviewed original research, review, or commentary article describing or discussing the use of social media in support of infectious disease surveillance (within the broad context of disease detection, prevention, and control) was considered relevant. Abstracts describing the use of social media within the context of educational or risk communication campaigns or strategies and those published in languages other than English, Spanish, or French were excluded due to their irrelevance to the scope of the review and limited resources for translation, respectively. Conflicts between reviewers were resolved by consensus or with the assistance of the corresponding author, when required. A list of all relevant articles identified at the relevance screening level was shared with the members of the Advisory Committee to identify if any potentially relevant citations were missed.

Data Characterization and Extraction

The full papers of relevant abstracts were procured and subsequently assessed by one reviewer (KR) to confirm their relevance. To ensure the accuracy of the data characterization, a random subsample of 19 articles was also independently reviewed by a second reviewer: TB (n=10), JF (n=9), MP (n=5). At this stage, the data characterization and extraction were limited to articles investigating or discussing the review question within the context of infectious disease. An a priori developed data characterization and extraction form consisted of 20 closed (n=15 questions) and open (n=5) questions. The closed questions captured the article type and format, sector and targeted audience, definitions of social media (if reported), study/surveillance/jurisdiction objectives, type of social media and surveillance method description, investigation of comparison and/or accuracy of social media versus other surveillance systems, and reported strengths and challenges associated with social media–based surveillance. Conflicts between reviewers were resolved by consensus or with the assistance of the corresponding author, when required. Data extracted from primary research articles were downloaded as MS Excel spreadsheets, summarized, and charted using narrative synthesis, tables, and figures.

Thematic Analysis

We conducted a thematic analysis of all identified review and case study articles (n=19) to determine the important characteristics, considerations, and challenges regarding the use of social media for infectious disease surveillance. Thematic analysis is a method of qualitative synthesis that involves the identification of key and recurrent themes and concepts from a body of literature [14]. The analysis was conducted by 2 independent reviewers (AR and IY) using an inductively developed form and code list (see Multimedia Appendix 1). The form and codes were informed by discussions from the workshop about the use of social media for disease surveillance and from reviewing a sample of five relevant articles. Both reviewers independently coded all documents and met periodically to compare and discuss their findings. After completion of coding, the 2 reviewers discussed and consolidated their results, then developed overall themes by grouping and consolidating codes that represented similar concepts.

Results

Search Strategy and Study Selection

The citation flow through various stages of the scoping review is shown in Figure 1. From 683 citations screened for relevance, 101 were considered potentially relevant and obtained as full articles.

Data Characterization and Extraction

During data characterization and extraction, 32 primary research articles and 19 reviews and case studies were identified as relevant (Figure 1). The data characteristics of 32 relevant primary research articles are displayed in Table 1, and the full list of relevant articles from data characterization and extraction is available in Multimedia Appendix 1.

Most relevant citations were peer-reviewed journal articles (29/32, 91%) published in 2010 and 2011 (28/32, 88%) and reported the use of a Google program (17/32, 53%, eg, Google Trends, Flu Trends, or Insights for Search) for surveillance of influenza (23/32, 72%) (Table 1 and Figure 2). Only four primary research articles investigated social media in the context of foodborne disease or gastroenteritis (Table 1). None of the articles provided a definition for social media. However, two articles referred to the term "infodemiology", which is defined as "the science of distribution and determinants of information in an electronic medium, specifically the Internet, or in a population, with the ultimate aim to inform public health and public policy" [1]. Use of infodemiology data for surveillance has been called "infoveillance" [1] or "digital disease detection" [15].

Most authors (21/32 articles, 66%) reported that the social media–based surveillance had good correlation when compared to an existing surveillance program (Table 2). The most commonly reported strengths of social media surveillance programs included their effectiveness (21/32, 66%) and rapid detection of disease trends (21/32, 66%). The most commonly reported weaknesses were the potential for false positive (16/32, 50%) and false negative (11/32, 34%) results (Table 2). Most authors (24/32, 75%) recommended that social media programs should primarily be used to support existing surveillance programs (Table 2).

Figure 1. Scoping review flow chart.
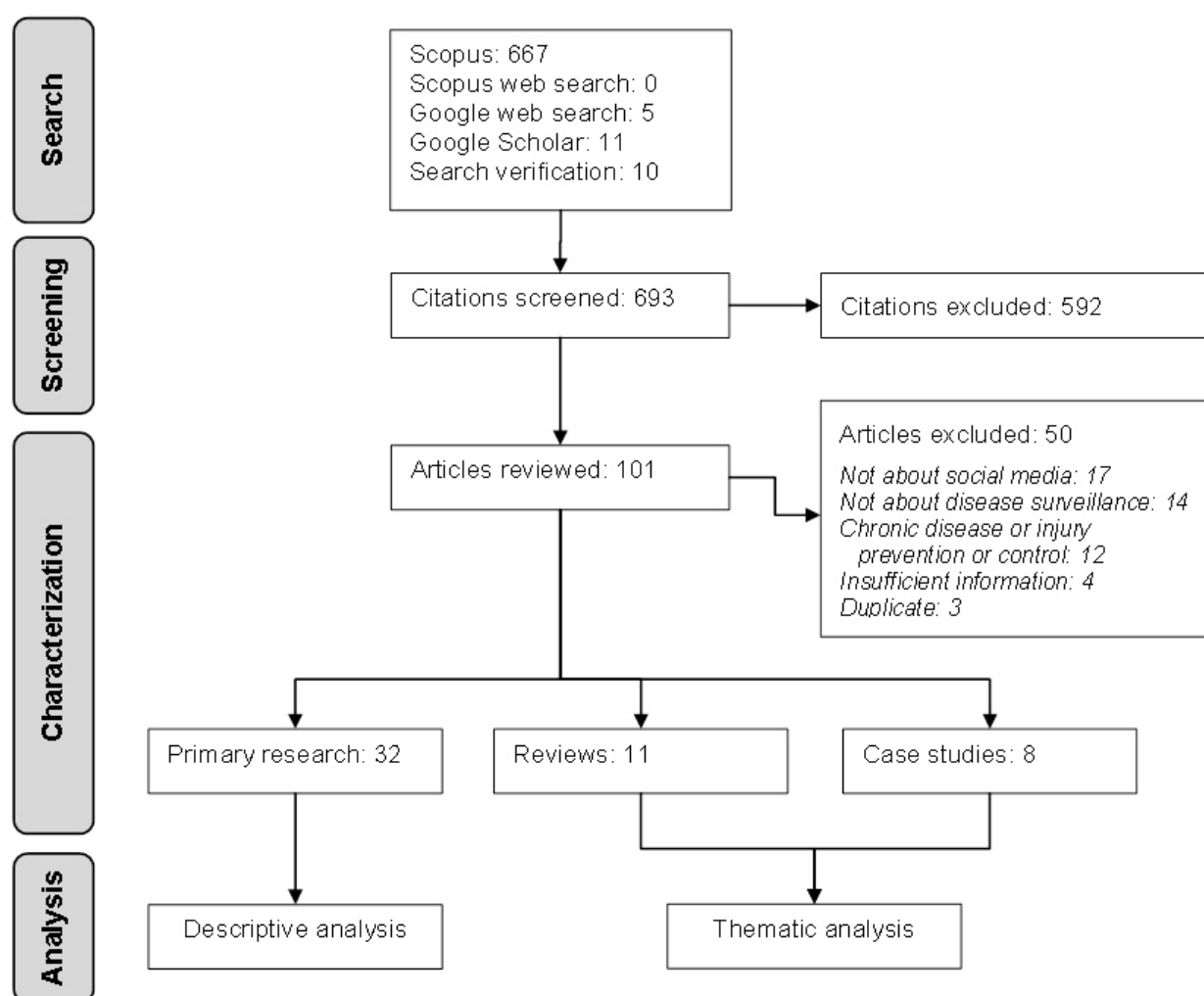
Table 1. Characteristics of 32 primary research articles investigating the use of social media for infectious disease surveillance published from 2002-2011.

Question			No.	%
Document type				
	Peer-reviewed journal article		29	90.6
	Book chapter		1	3.1
	Workshop report		1	3.1
	Conference proceedings abstract		1	3.1
Year of publication				
	2011		13	40.6
	2010		15	46.9
	2006-2009		5	15.6
Target audience [a]				
	Researchers and academics		29	90.6
	Practitioners, clinicians, or service providers		7	21.9
	Policy and decision makers		1	3.1
Jurisdictional level of surveillance				
	National		26	81.3
		USA	12	37.5
		Canada	2	6.3
		China	2	6.3
		UK	2	6.3
		Other[b]	8	25.0
	International		6	18.8
Social media program investigated [a]				
	Google		17	53.1
		Google Trends	5	15.6
		Google Flu Trends	4	12.5
		Google Search	4	12.5
		Google Insights for Search	3	9.4
		Google AdSense	1	3.1
	Twitter		10	31.3
	Yahoo		2	6.3
		Yahoo Search	1	3.1
		Yahoo Knowledge public health forums	1	3.1
	Other search engine[c]		3	9.4
	Blogs or Web forum		2	6.3
Infectious disease investigated [a]				
	Influenza (seasonal and highly pathogenic)		23	71.9
	Foodborne disease / gastroenteritis		4	12.5
	Dengue		3	9.4
	HIV/AIDS		2	6.3
	Other[d]		4	12.5

XSL•FO
RenderX

[a]Multiple answers allowed per article (ie, percentages do not add to 100%).

[b]Other countries included Australia, Brazil, France, Germany, Japan, Spain, Sweden, and Taiwan.

[c]Included Baidu (n=2) and Vardguiden (n=1).

[d]Other diseases included scarlet fever, tuberculosis, Lyme disease, methicillin-resistant Staphylococcus aureus, chickenpox, and ophthalmologic conditions.

Table 2. Characteristics of social media programs for infectious disease surveillance as reported in 32 primary research articles published from 2002-2011.

Question		No.	%
Accuracy of the social media program compared to an existing surveillance program [a]			
	The compared systems showed good correlation	21	65.6
	The social media program was more accurate	2	6.3
	The existing program was more accurate	2	6.3
	Not reported	1	3.1
	No comparison conducted	7	21.9
Reported strengths of social media programs for infectious disease surveillance [b]			
	Effective	21	65.6
	Faster response/detection	21	65.6
	Cost-effective	9	28.1
	Easy to access	7	21.9
	User-friendly	4	12.5
	Unique/global population as data source	4	12.5
	Less resource intensive	3	9.4
	Flexible	3	9.4
Reported weaknesses of social media programs for infectious disease surveillance [b]			
	Potential for false positives (eg, increased searching due to media reporting)	16	50.0
	Potential for false negatives (eg, social media users might not represent general public)	11	34.4
	Variability in the function of different social media tools	4	12.5
	User information privacy concerns	2	6.3
	Sufficient skills and timely use needed to be effective	1	3.1
Reported recommendations for the use of social media programs for infectious disease surveillance [b]			
	Should primarily support existing programs	24	75.0
	Should be used in the future when the methods are better validated and evaluated	3	9.4
	Should be used as a proxy for existing programs or when no traditional surveillance program exists	3	9.4
	Not reported	3	9.4

[a]One article had two responses based on differences in program performance for different diseases investigated.

[b]Multiple answers allowed per article (ie, percentages do not add to 100%).

Figure 2. Distribution of published primary research investigating the use of social media programs for infectious disease surveillance from 2002-2011 (N=39 in this graph because some articles investigated more than one disease or used more than one social media program).
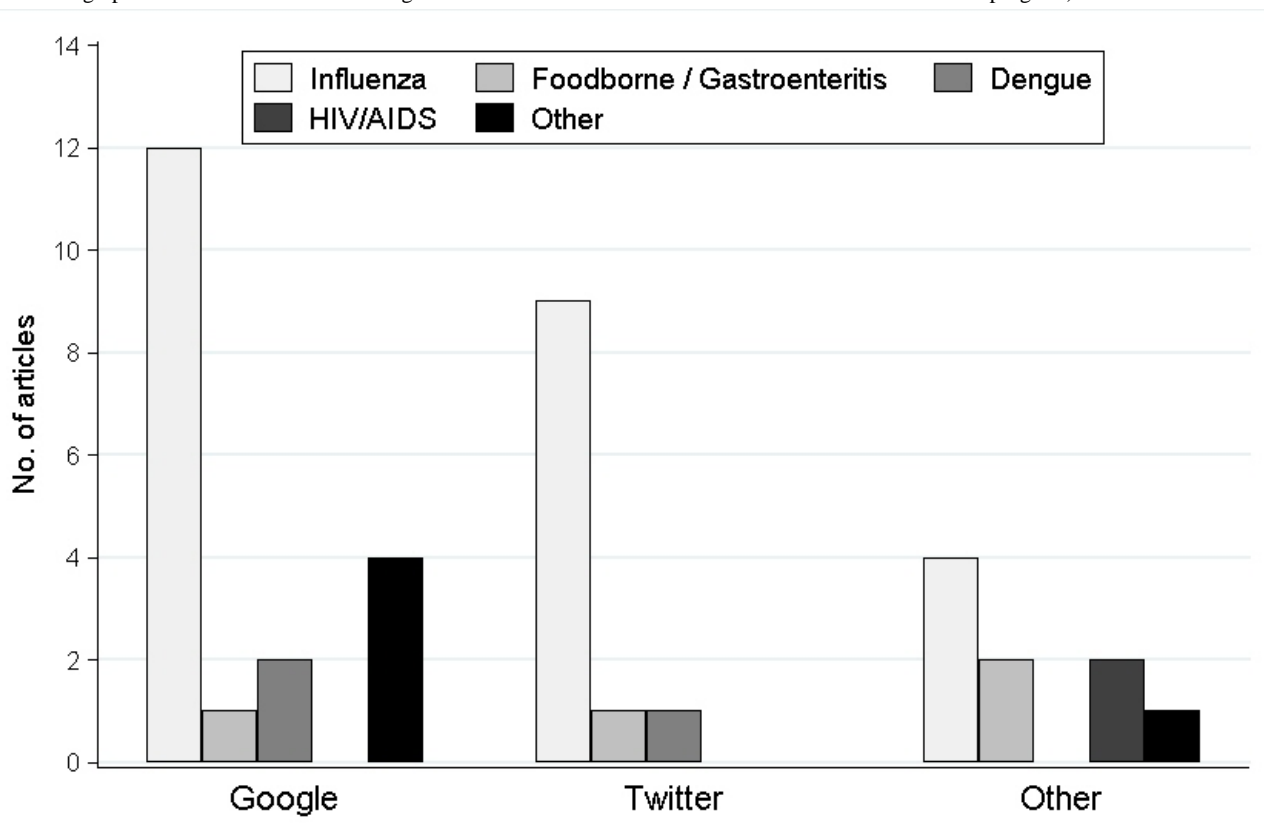
Thematic Analysis

Four thematic areas were identified as key characteristics of social media-based surveillance in the context of infectious disease (Figure 3). The first theme relates to the methodological aspects of the programs. In general, a variety of ontologies and search algorithms is used to synthesize and filter unstructured information from a variety of Web-based sources [15-21]. These sources can include news aggregates (eg, ProMed-mail), social media platforms (eg, Twitter), blogs, and search engine queries (eg, Google). Data sources can be characterized further as supply-based (eg, blogs and social media) or demand-based (eg, search behaviors) [18]. The overall principle behind these programs is that they aim to make sense of the public's "collective intelligence" for purposes of early detection and effective control of infectious disease [15,18].

A second identified theme was the necessary capacity for developing these programs in practice. A multidisciplinary and multijurisdictional approach is needed to allow adequate data collection, exchange, and evaluation and communication across multiple jurisdictions and wide geographical areas [18,20,21]. Social media programs can allow international networks of food safety, public health, and other professionals to communicate via virtual networks, which can facilitate collaborations and support public health response infrastructure [19-22]. One example is virtual situation rooms using a three-dimensional interface, where public health professionals can collaborate and discuss surveillance data in real-time [20]. However, government and public health officials must be adequately trained and skilled

in order to utilize these tools for disease surveillance in an effective and timely way [15,21].

Several advantages were frequently pointed out regarding the social media-based surveillance programs for infectious disease. First, the identification of disease trends in real-time, which can contribute to rapid outbreak detection and response [15,17-21]. In addition, they tend to be openly accessible to the public, be low cost or free, have a familiar and user-friendly interface, and have potential applications and benefits for multiple end-users (eg, public health officials, media, and travellers) [15,17-21,23,24]. In confirmation with our analysis of primary research articles (Table 2), these programs are primarily recommended as supplementary applications to existing surveillance programs, or as Madoff et al [25] note: "another tool in the surveillance toolbox."

Finally, multiple challenges to the use of social media programs for infectious disease surveillance were identified. One of the most important challenges relates to the validity and reliability of the data analysis. For example, several authors discussed the need to properly filter out background noise (eg, people searching out of curiosity rather than illness) to ensure that the surveillance data reflect actual disease trends and are not a result of heightened media exposure or other biases [15,17,19,21,23,25,26]. In addition, there are still certain segments of the population that do not regularly use the Internet or social media programs, particularly in developing countries, so the users of these programs may not accurately represent the general population [20,24,25,27]. Another challenge relates to the ownership of the data and the issues surrounding user

information privacy and confidentiality [21,27]. Most authors agreed that more evaluation, validation, and development of

these programs is needed before they should be widely used in practice [17,20,25,26,28].

Figure 3. Key characteristics of social media programs for infectious disease surveillance.
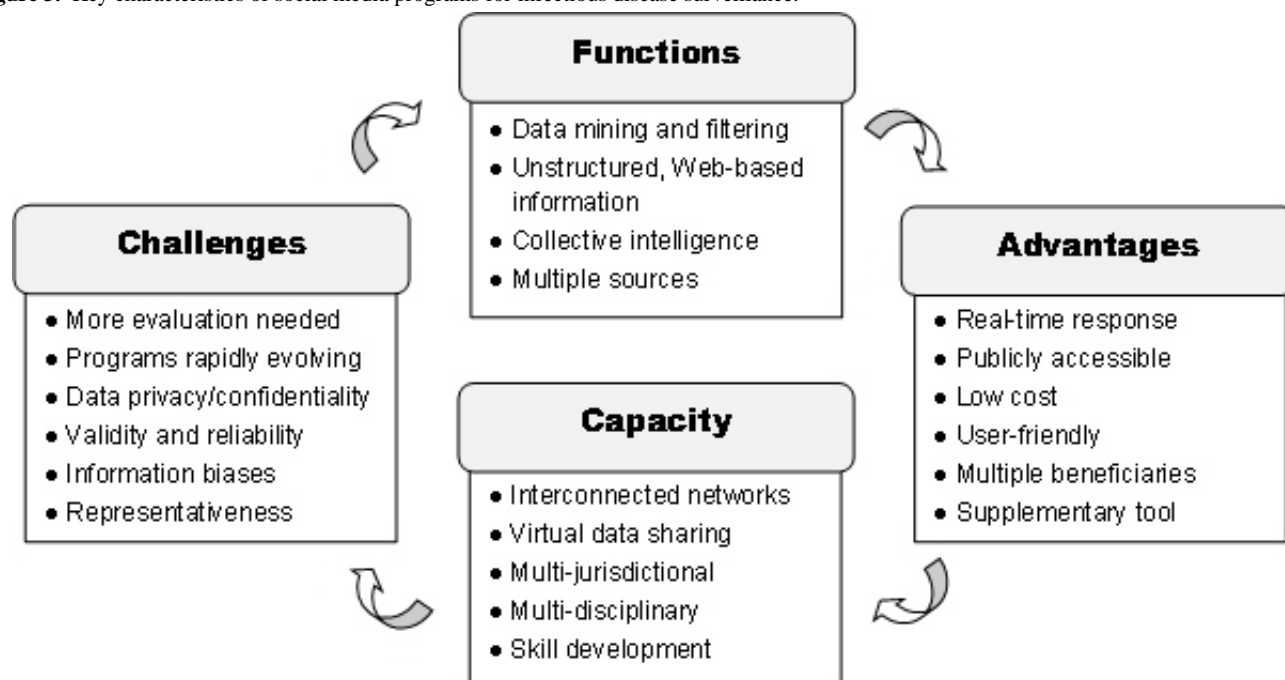
Discussion

Overview

The relevant research identified by this scoping review included a total of 51 articles, most of which were published since 2010 and investigated applications for enhancing influenza surveillance. This low to moderate yield of research activity was expected, as neither Web searches nor social media were developed with the objective of disease surveillance in mind and they are relatively recent phenomenon. As is frequently the case with innovation, new uses of existing tools are driven by necessity and/or opportunity.

Experimentation with search queries and social media for disease surveillance appears to reflect the chronological availability of new tools and the concurrent disease surveillance challenges, as well as the development of data mining and machine learning techniques. This may explain, in part, why the most common approach was to use Google-related search tools, as their chronological development preceded other social media tools, as well as Google's more global scope in availability for application.

Chronology of Development

As the use of Web searches to obtain health information became commonplace, researchers turned from following the number of people searching for health information, to looking at whether the frequency of searches on particular subjects harbored useful data, such as clues to disease outbreaks. The earliest article identified by this scoping review by Eysenbach [1] was published before search query data were widely available. Eysenbach devised a clever method to circumvent this restriction and acquired data on searches related to influenza through a strategic combination of bids for targeted Google keywords and

placement of an influenza-related advertisement. He then developed a model for detecting influenza outbreaks in Canada based on changes in Canadians' searches for information on influenza. When evaluated against the gold standard for influenza surveillance (reports by sentinel physicians of clinical encounters with influenza-like illness), the model proved to be more timely, accurate, and inexpensive [1]. The benefits reported in this earliest publication reflect the main benefits of social media-facilitated disease surveillance identified in the literature included in the scoping review.

The infectious disease most commonly evaluated using social media surveillance techniques was influenza, which is not surprising as these tools became available during a period of heightened sensitivity to the threat of a global pandemic following outbreaks of influenza H5N1 in 1997 and SARS in 2002. A study by Polgreen et al [3] found that the frequency of searches for influenza had predictive potential in the United States, looking at data over a longer time period (2004 to 2008) and using a different search engine (Yahoo!) than Eysenbach [1]. They were able to predict an increase in positive cultures for influenza 1-3 weeks before the increase occurred ($P<.001$) and an increase in mortality attributable to pneumonia and influenza up to 5 weeks in advance ($P<.001$). Two of the authors were employees at Yahoo!, which accounts for their access to search data [3].

In 2009, a letter authored by employees of Google and the Centers for Disease Control and Prevention (CDC) published in *Nature* described a large-scale effort to use Google search queries to track influenza [2]. A model was created based on the top 45 queries most correlated with CDC data on influenza-like illness. It consistently estimated the level of weekly influenza activity in each region of the United States with a 1-day reporting lag, which was 1-2 weeks ahead of

reports by the CDC's US Influenza Sentinel Provider Surveillance Network. Perhaps most importantly, results were made freely available online at Google Flu Trends website. This methodology was extended to Google Dengue Trends and was then generalized as Google Correlate, which allows users to enter their own search terms or time series data to find other terms that have a similar pattern of activity.

Pelat et al [29] demonstrated that using search queries for disease detection also functioned in another language (ie, French) and could be applied to other diseases (ie, gastroenteritis and chicken pox). The symptom of gastroenteritis, used as an indicator of foodborne illness, is of particular significance due to the difficulty in detecting foodborne illness in a timely manner. Whereas there is a lag of 1-2 weeks in tracking influenza, most foodborne disease outbreaks are not detected for several months after they occur, by which time the outbreak and opportunity for intervention are virtually over. A Chinese study by Zhou et al in 2010 [30] used both Baidu search queries and Baidu news articles to track infectious diseases including dysentery. They were able to reduce the distorting effect of disease-related news reports by using a combination of search frequency data and news count data. Surveillance reports from this effort were published 10-40 days ahead of the release of official reports from the Chinese government CDC.

Publications on the use of Twitter first appeared in 2010 and followed a similar pattern to the use of Internet search queries: they predominantly dealt with influenza (Figure 2) and ranged from content analysis of Twitter messages (tweets) related to the H1N1 outbreak [31,32] to demonstrating that tweets could accurately track an outbreak [31]. After analyzing over 570 million tweets, Culotta (2010) [33] concluded that "even extremely simple methods can result in quite accurate models" of influenza rates. Models are improved through judicious selection of keywords to track and by devising better methods to filter spurious tweets through natural language processing [34]. Content analysis of German tweets was also conducted for a number of diseases including influenza, norovirus, and salmonella [35].

Geolocation

In addition to determining when an outbreak is occurring, it would be useful to know where it is occurring. Although geolocation was not targeted for evaluation in the scoping review, researchers included it as a possible use. The general physical location of a search query's origin can often be identified from its associated Internet protocol (IP) address [2]. Although Twitter has an optional geolocation feature, a recent publication found that the prevalence of tweets with geolocation data was only 2%; however, city and state could be determined for 17% of user profiles using a simple text-matching approach [36]. Agreement between GPS data and text-matching was high (88%), as was the correlation between the number of geolocated tweets and state populations in the United States (ie, geolocated tweets were proportional to the state population) [36].

Two mapping systems were launched in 2006, BioCaster [16] and HealthMap [23], that monitor news feeds in multiple languages to provide real-time intelligence on emerging diseases around the world. Sources including news media, discussion

sites such as ProMED-mail, and official reports of international organizations. HealthMap's interface provides a means of organizing unstructured information based on geography, time, and infectious disease agent. HealthMap currently invites user input on missing outbreaks and includes a feature that solicits user contributions on influenza illness symptoms called "flu near you".

The first articles describing mapping of tweets appeared in 2011. Signorini et al [31] created a Google map continuously updated with selected tweets to provide a real-time view of influenza-related public sentiment. Gomide et al [37] proposed a method for dengue surveillance in Brazil using four dimensions of Twitter data—volume, location, time, and content—in which they looked at the proportion of tweets expressing personal experience with dengue. Spatio-temporal analysis of dengue to detect clusters would enable government agencies to concentrate efforts in the right place at the right time.

Participatory Surveillance

The potential of social media for epidemiology goes beyond the passive generation of new data streams from people, animals, food, or other sensors, and their movements. People can be actively involved in, or even instigate, epidemiological investigations. For example, postings in a Web forum about ill participants following a bike race in 2007 prompted the organizers to notify local public health authorities [38]. Messages and photos on the Web forum provided contextual clues as to the source (mud) of the outbreak (laboratory confirmed *Campylobacter jejuni*) that might have otherwise been missed, and an online questionnaire hastened the outbreak investigation [38].

Another example occurred in February 2011, when an Internet entrepreneur became sick after attending an international conference and posted a status update on Facebook [39]. Within a week, 80 other participants from around the world had self-identified and arrived at a potential diagnosis of legionellosis. The officer assigned to the case from the CDC joined the Facebook page to read the history of the outbreak and recommended appropriate diagnostic tests. This is an extreme example of participatory epidemiology whereby the investigation was initiated by those affected and epidemiologists were invited to participate. Social media is a breakthrough technology because it reduces the cost and difficulty of forming and working in groups, making it possible for loosely affiliated people to accomplish things that once were only possible through formal organizations [40].

Potential for Adoption

Official reports by governments and international organizations were the primary source of disease intelligence during the 20th century. Unofficial reports were first taken into consideration by the moderated mailing list ProMED-mail, which was launched in 1994 [15]. Detection and investigation of "rumors" from news feeds and websites formed the basis of the Global Public Health Intelligence Network in 1997: a joint project of the Public Health Agency of Canada and the World Health Organization [1,15]. These examples set a precedent for

the adoption of search queries and social media as a supplement to existing surveillance activities, in keeping with the reported recommendations of the scoping review (Table 2). Adoption, even as supporting systems for existing surveillance, will entail a high level of familiarity with the tools and collaboration across organizations and jurisdictions.

There is a growing body of evidence for the utility and accuracy of search queries in tracking diseases. The textual content of a tweet, however, differentiates it from search query data and may provide additional useful and timely information [31]. Computers can learn to distinguish useful messages based on word associations providing an automated method to deal with millions of tweets, using tools such as the Support Vector Machine (SVM)-based classifier [31,34]. The potential for false positives and false negatives was identified as one of the most commonly reported weaknesses by this scoping review (Table 2). One of the primary challenges is to refine the data signal by reducing surrounding noise. Further developments in digital disease surveillance have the potential to improve sensitivity and specificity: passively through advances in machine learning and actively through engagement of users.

Most of the identified research to date is associated with using Google search queries to detect seasonal or pandemic influenza days to weeks in advance of existing surveillance programs, but there are other promising areas for improvement. Just as influenza can be transported around the world in a matter of hours, our increasingly complex global food-supply chain presents a growing challenge to governments attempting to ensure a safe food supply in the face of dwindling budgets. Foodborne outbreaks can be notoriously difficult to detect as they can be widely distributed geographically and may be due to an ingredient that is found in a number of foods. Foodborne illness is also vastly underreported since most people who are affected do not seek medical attention nor receive laboratory confirmation of the causative agent (necessary steps to trigger declaration of an outbreak). Newkirk et al [41] make the case for using real-time data from social media to bypass significant delays in traditional foodborne surveillance activities, estimating a potential savings of 5-19 days in the reporting timeline for salmonellosis.

Limitations

A potential limitation of this review is that only one electronic database was used to identify literature; however, we believe that our search verification strategy helped to limit this potential bias and are confident that the review was robust, results are accurate, and all relevant articles published during the study period were included. Another limitation of this review is the potential bias introduced by having only 1 reviewer extract data from the primary research articles during the data characterization and extraction step. However, we are confident that these results are accurate given that only minor conflicts were identified among the sample of articles verified by a second reviewer. In addition, many of our key results and conclusions correspond to and build upon those of other recently published reviews in this area [41,42].

Conclusions

The use of search queries and social media for disease surveillance are relatively recent phenomena. Both the tools themselves and the methodologies for exploiting them are evolving over time. The growing evidence base regarding the utility of social media for disease surveillance will hopefully encourage academia, industry, the public service, and international organizations to consider social media in a serious light, particularly as a means of engagement rather than just disseminating information. While their accuracy, speed, and cost compare favorably with existing surveillance systems, the primary challenge is to refine the data signal by reducing surrounding noise. Further developments in digital disease surveillance have the potential to improve sensitivity and specificity: passively through advances in machine learning and actively through engagement of users. Although learning to use and adapt these new tools will take some time and effort, the greater challenge will be the multilevel collaboration among local, regional, national, and international authorities that will be required to use them most effectively.

Authors' Contributions

Dr. Theresa Bernardo led the manuscript development and provided expertise in social media and disease surveillance, Dr. Andrijana Rajic led in the conduct of the scoping review methodology, and Dr. Julie Funk provided content expertise in disease surveillance. Dr. Andrijana Rajic and Dr. Ian Young conducted the thematic analysis, Mai Pham conducted the searches, and Katie Robiadek led data characterization and extraction and carried out Google searches. All authors evaluated the studies and contributed to the manuscript development and revisions.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Supplementary material.

[PDF File (Adobe PDF File), 316KB - jmir_v15i7e147_app1.pdf]

References

1. Eysenbach G. Infodemiology: tracking flu-related searches on the web for syndromic surveillance. AMIA Annu Symp Proc 2006:244-248 [FREE Full text] [Medline: 17238340]

2. 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]

3. Polgreen PM, Chen Y, Pennock DM, Nelson FD. Using internet searches for influenza surveillance. Clin Infect Dis 2008 Dec 1;47(11):1443-1448 [FREE Full text] [doi: 10.1086/593098] [Medline: 18954267]

4. Sargeant JM, Torrence ME, Rajić A, O'Connor AM, Williams J. Methodological quality assessment of review articles evaluating interventions to improve microbial food safety. Foodborne Pathog Dis 2006;3(4):447-456. [doi: 10.1089/fpd.2006.3.447] [Medline: 17199527]

5. Waddell L, Rajić A, Sargeant J, Parker S, Deckert A, McEwen S. The methodological soundness of literature reviews addressing three potential zoonotic public health issues. Zoonoses Public Health 2009 Nov;56(9-10):477-489. [doi: 10.1111/j.1863-2378.2008.01194.x] [Medline: 19175572]

6. Arksey H, O'Malley L. Scoping studies: towards a methodological framework. International Journal of Social Research Methodology 2005 Feb;8(1):19-32. [doi: 10.1080/1364557032000119616]

7. Davis K, Drey N, Gould D. What are scoping studies? A review of the nursing literature. Int J Nurs Stud 2009 Oct;46(10):1386-1400. [doi: 10.1016/j.ijnurstu.2009.02.010] [Medline: 19328488]

8. Farrar AM. The application of research synthesis methods for evaluating primary research on Salmonella in broiler chickens. ProQuest Dissertations and Theses 2009.

9. Ilic S, Rajić A, Britton CJ, Grasso E, Wilkins W, Totton S, et al. A scoping study characterizing prevalence, risk factor and intervention research, published between 1990 and 2010, for microbial hazards in leafy green vegetables. Food Control 2012 Jan;23(1):7-19. [doi: 10.1016/j.foodcont.2011.06.027]

10. Tuševljak N, Rajić A, Waddell L, Dutil L, Cernicchiaro N, Greig J, et al. Prevalence of zoonotic bacteria in wild and farmed aquatic species and seafood: a scoping study, systematic review, and meta-analysis of published research. Foodborne Pathog Dis 2012 Jun;9(6):487-497. [doi: 10.1089/fpd.2011.1063] [Medline: 22571642]

11. Michigan State University. Food Safety Biosurveillance Workshop. 2011. URL: https://fsb.cvm.msu.edu/ [accessed 2013-07-11] [WebCite Cache ID 6I2ZpqrQg]

12. Kaplan AM, Haenlein M. Users of the world, unite! The challenges and opportunities of Social Media. Bus Horiz 2010;53(1):59-68. [doi: 10.1016/j.bushor.2009.09.003]

13. MedicineNet. Definition of Disease surveillance. 2012. URL: http://www.medterms.com/script/main/art.asp?articlekey=26702 [accessed 2013-04-25] [WebCite Cache ID 6G92xcFKL]

14. Dixon-Woods M, Agarwal S, Jones D, Young B, Sutton A. Synthesising qualitative and quantitative evidence: a review of possible methods. J Health Serv Res Policy 2005 Jan;10(1):45-53. [Medline: 15667704]

15. Brownstein JS, Freifeld CC, Madoff LC. Digital disease detection--harnessing the Web for public health surveillance. N Engl J Med 2009 May 21;360(21):2153-5, 2157 [FREE Full text] [doi: 10.1056/NEJMp0900702] [Medline: 19423867]

16. Collier N, Doan S, Kawazoe A, Goodwin RM, Conway M, Tateno Y, et al. BioCaster: detecting public health rumors with a Web-based text mining system. Bioinformatics 2008 Dec 15;24(24):2940-2941 [FREE Full text] [doi: 10.1093/bioinformatics/btn534] [Medline: 18922806]

17. 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]

18. 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]

19. Greene M. Epidemiological monitoring for emerging infectious diseases. Sensors, and Command, Control, Communications, and Intelligence (C3I) Technologies for Homeland Security and Homeland Defense IX; 5 April 2010 through 9 April 2010; Orlando, FL 2010. [doi: 10.1117/12.849351]

20. Kamel Boulos MN, Sanfilippo AP, Corley CD, Wheeler S. Social Web mining and exploitation for serious applications: Technosocial Predictive Analytics and related technologies for public health, environmental and national security surveillance. Comput Methods Programs Biomed 2010 Oct;100(1):16-23. [doi: 10.1016/j.cmpb.2010.02.007] [Medline: 20236725]

21. Khan AS, Fleischauer A, Casani J, Groseclose SL. The next public health revolution: public health information fusion and social networks. Am J Public Health 2010 Jul;100(7):1237-1242. [doi: 10.2105/AJPH.2009.180489] [Medline: 20530760]

22. Lopes LF, Silva FAB, Couto F, Zamite J, Ferreira H, Sousa C, et al. Epidemic marketplace: An information management system for epidemiological data. Lect Notes Comput Sci 2010;6266:31-44. [doi: 10.1007/978-3-642-15020-3_3]

XSL•FO
RenderX

23. Brownstein JS, Freifeld CC, Reis BY, Mandl KD. Surveillance Sans Frontières: Internet-based emerging infectious disease intelligence and the HealthMap project. PLoS Med 2008 Jul 8;5(7):e151 [FREE Full text] [doi: 10.1371/journal.pmed.0050151] [Medline: 18613747]

24. Schein R, Wilson KKeelan J. Literature Review on Effectiveness of the Use of Social Media: a report for Peel Public Health. 2010. URL: http://www.peelregion.ca/health/resources/pdf/socialmedia.pdf [accessed 2013-04-25] [WebCite Cache ID 6G92PIsCG]

25. Madoff LC, Fisman DN, Kass-Hout T. A new approach to monitoring dengue activity. PLoS Negl Trop Dis 2011 May;5(5):e1215 [FREE Full text] [doi: 10.1371/journal.pntd.0001215] [Medline: 21647309]

26. Dreesman J, Denecke K. Challenges for signal generation from medical social media data. Stud Health Technol Inform 2011;169:639-643. [Medline: 21893826]

27. Eke PI. Using social media for research and public health surveillance. J Dent Res 2011 Sep;90(9):1045-1046. [doi: 10.1177/0022034511415277] [Medline: 21768305]

28. Cook AR, Chen MI, Pin Lin RT. Internet search limitations and pandemic influenza, Singapore. Emerg Infect Dis 2010 Oct;16(10):1647-1649 [FREE Full text] [doi: 10.3201/eid1610.100840] [Medline: 20875307]

29. Pelat C, Turbelin C, Bar-Hen A, Flahault A, Valleron A. More diseases tracked by using Google Trends. Emerg Infect Dis 2009 Aug;15(8):1327-1328 [FREE Full text] [doi: 10.3201/eid1508.090299] [Medline: 19751610]

30. Zhou X, Shen H. Notifiable infectious disease surveillance with data collected by search engine. J Zhejiang Univ - Sci C 2010 Apr 17;11(4):241-248. [doi: 10.1631/jzus.C0910371]

31. Signorini A, Segre AM, Polgreen PM. The use of Twitter to track levels of disease activity and public concern in the U.S. during the influenza A H1N1 pandemic. PLoS One 2011;6(5):e19467 [FREE Full text] [doi: 10.1371/journal.pone.0019467] [Medline: 21573238]

32. 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]

33. Culotta A. Detecting influenza outbreaks by analyzing Twitter messages. 2010. URL: http://www2.southeastern.edu/Academics/Faculty/aculotta/pubs/culotta10detecting.pdf [accessed 2013-04-25] [WebCite Cache ID 6G985XHKt]

34. Aramaki E, Maskawa S, Morita M. Twitter Catches the Flu: Detecting influenza epidemics using Twitter. Edinburgh, Scotland: Association for Computational Linguistics; 2011.

35. Krieck M, Dreesman J, Otrusina L, Denecke K. A New Age of Public Health: Identifying Disease Outbreaks by Analyzing Tweets. In: Health Web Science Workshop, ACM Web Science 2011 3rd International Conference on Web Science. 2011 Presented at: 3rd International Conference on Web Science; June 14-17, 2011; Koblenz, Germany.

36. Burton SH, Tanner KW, Giraud-Carrier CG, West JH, Barnes MD. "Right time, right place" health communication on Twitter: value and accuracy of location information. J Med Internet Res 2012;14(6):e156 [FREE Full text] [doi: 10.2196/jmir.2121] [Medline: 23154246]

37. Gomide J, Veloso A, Meira W, Almeida V, Benevenuto F, Ferraz F, et al. Dengue surveillance based on a computational model of spatio-temporal locality of Twitter. In: Proceedings of the 3rd International Conference on Web Science. 2011 Presented at: 3rd International Conference on Web Science; June 14-17, 2011; Koblenz, Germany.

38. Chester TLS, Taylor M, Sandhu J, Forsting S, Ellis A, Stirling R, et al. Use of a Web Forum and an Online Questionnaire in the Detection and Investigation of an Outbreak. Online Journal of Public Health Informatics 2011;3(1).

39. Garrity B. The New York Times. 2011 Jun 13. Social Media Join Toolkit for Hunters of Disease URL: http://www.nytimes.com/2011/06/14/health/research/14social.html?pagewanted=all&_r=0 [WebCite Cache ID 6GylEGxOn]

40. Shirky C. Here comes everybody: the power of organizing without organizations. New York, NY: The Penguin Press; 2008.

41. Newkirk RW, Bender JB, Hedberg CW. The potential capability of social media as a component of food safety and food terrorism surveillance systems. Foodborne Pathog Dis 2012 Feb;9(2):120-124. [doi: 10.1089/fpd.2011.0990] [Medline: 22217109]

42. Guy S, Ratzki-Leewing A, Bahati R, Gwadry-Sridhar F. Social media: A systematic review to understand the evidence and application in infodemiology. Lect Notes Inst Comput Sci Soc Informatics Telecommun Eng 2012;91:1-8. [doi: 10.1007/978-3-642-29262-0_1]

Abbreviations

CDC: Centers for Disease Control and Prevention
SARS: Severe Acute Respiratory Syndrome
SVM: Support Vector Machine

XSL•FO
RenderX

XSL•FO
RenderX

Original Paper

Long-Term Doctor-Patient Relationships: Patient Perspective From Online Reviews

Alissa Detz[1], MD; Andrea López[2], B.Sci; Urmimala Sarkar[2], MD, MPH

[1]UCLA Division of General Internal Medicine and Health Services Research, University of California, Los Angeles, Los Angeles, CA, United States
[2]Center for Vulnerable Populations, Division of General Internal Medicine, University of California, San Francisco, San Francisco, CA, United States

Corresponding Author:
Urmimala Sarkar, MD, MPH
Center for Vulnerable Populations
Division of General Internal Medicine
University of California, San Francisco
1001 Potrero Ave
Bldg 10, FL 3, Ward 13
San Francisco, CA, 94110
United States
Phone: 1 415 206 6962
Fax: 1 415 206 5586
Email: usarkar@medsfgh.ucsf.edu

Abstract

Background: Continuity of patient care is one of the cornerstones of primary care.

Objective: To examine publicly available, Internet-based reviews of adult primary care physicians, specifically written by patients who report long-term relationships with their physicians.

Methods: This substudy was nested within a larger qualitative content analysis of online physician ratings. We focused on reviews reflecting an established patient-physician relationship, that is, those seeing their physicians for at least 1 year.

Results: Of the 712 Internet reviews of primary care physicians, 93 reviews (13.1%) were from patients that self-identified as having a long-term relationship with their physician, 11 reviews (1.5%) commented on a first-time visit to a physician, and the remainder of reviews (85.4%) did not specify the amount of time with their physician. Analysis revealed six overarching domains: (1) personality traits or descriptors of the physician, (2) technical competence, (3) communication, (4) access to physician, (5) office staff/environment, and (6) coordination of care.

Conclusions: Our analysis shows that patients who have been with their physician for at least 1 year write positive reviews on public websites and focus on physician attributes.

Introduction

In the United States, recent health reform legislation has increasingly emphasized patient-centered care and patient satisfaction within primary care. Patients often have a choice when selecting a primary care physician. Therefore, patient reviews of their experiences may influence choice of physician as well as physician practices.

Continuity of patient care is one of the cornerstones of primary care [1]. Previous research indicates that both patients and physicians value this aspect of outpatient medical care [2-4].

Moreover, continuity of care is associated with improved management of chronic disease, increased administration of preventative health services, and fewer emergency department visits and hospitalizations [5-11].

The patient-centered medical home is a model of providing primary care defined by management of a population of patients rather than provision of care during periodic primary care visits. This model emphasizes patient-centeredness, accessibility, and comprehensive and coordinated care with a focus on patient safety and quality. The medical home model has been increasingly promoted as a means to improve primary care in

the United States and emphasizes continuity between patient and provider as a core component [12-14].

Despite the importance of continuity, there are few studies dedicated to defining what factors are important for establishing and maintaining a relationship with a given physician over time. Since promoting continuity of care is an explicit goal in providing quality primary care, identifying factors that promote continuity is critical. In turn, this requires understanding patient perspectives on long-term relationships with primary care physicians.

Although Internet website reviews of physicians are controversial [15-17], they do provide unfiltered data regarding patient perceptions of health care. These reviews can complement existing studies on the patient-physician relationship. Traditional structured satisfaction surveys have been shown to perform differently across patient populations and may not capture the views of all patients [18-21]. Public websites allow individuals to review their physicians in an anonymous and unstructured format. Evaluating these publically available unstructured reviews may give us additional insight into what factors of the patient-doctor relationship are particularly important to patients. In this study, we examine publicly available, Internet-based reviews of adult primary care physicians, specifically written by patients who report long-term relationships with their physicians. We employed qualitative analysis to undercover themes within reviews of long-term patient-physician relationships.

Methods

Design

This substudy was nested within a larger qualitative content analysis of online physician ratings. The methods of the parent study are described in detail elsewhere [22]. The parent study was a qualitative content analysis of 712 online reviews from two publicly available rating websites (Yelp, a general rating site and RateMDs, a physician-rating website). For the parent study, we purposively sampled reviews of 445 primary care doctors (internists and family practitioners) from four geographically dispersed urban locations in the United States.

For this substudy, we focused on reviews reflecting an established patient-physician relationship. We chose this subset of reviews due to our interest in continuity of care. We defined long-term patients as those seeing their physicians for at least 1 year. There is a lack of consensus about what constitutes a long-term patient-physician relationship. Time frames are commonly defined by either number of visits or calendar time. We elected to use a 1-year time frame because previous investigators [23-26] and multiple Internet reviews in our dataset referenced this time frame.

Sampling

In the parent study, our search strategy was meant to mimic two popular ways of searching for ratings using the Internet: (1) using a search engine and (2) using a well-known general ratings site. First, to mimic a patient's approach, we utilized the popular Google search engine. When we entered the phrase "rate doctor" into Google.com, the first result was for the website RateMDs. As its name suggests, RateMDs exclusively rates physicians. Second, because we surmised that patients might search for physician ratings on a website they use for other types of consumer ratings, we selected the website Yelp. Our sampling strategy had two distinct levels because each physician could have multiple reviews. Each website first generates a list of physicians. Because the order in which doctors were listed on the website is nonrandom, we prespecified our sampling of physicians as follows: We selected 30 reviews of doctors appearing at the beginning of the search results list, 40 reviews of doctors appearing in the middle of the search results list, and 30 reviews of doctors appearing at the end of the search results list. Next, we purposively sampled the first three available reviews for each individual physician. We analyzed reviews that patients posted publicly. We de-identified physicians (the reviews' subjects) and identified overarching themes rather than focusing on individual performance. Moreover, the patients (review authors), who knowingly posted reviews publicly, did so with varying degrees of anonymity (true name vs Yelp username) and revealed differing amounts of personal data. For ethical reasons, we chose to de-identify review author data prior to analysis, even for individuals who designated their information as public. Utilizing the parent study, we extracted all patient reviews that referred to amount of time with their physician. Of 712 reviews, 3 patients specified a relationship with their physician of 1 year, 7 specified a relationship of 1-2 years, 74 specified a relationship of greater than 2 years, and 16 patients did not specify a number of years but implied a long-term relationship through their comments ("several years"). The remainder of reviewers did not specify length of time with a physician (Figure 1).

Qualitative Analysis

As explained in the parent study, we developed preliminary codes of all reviews by applying content analysis theory to a sample set of 50 reviews [27,28]. When developing our codes, we incorporated themes from the literature about factors in patient-physician encounters that impact patient satisfaction [29,30].

Two investigators independently coded 328 (46%) of the reviews, and the remainder of the reviews were coded by 1 investigator. Codes were created as new themes emerged and thematic saturation was achieved after 100 reviews. A total of 60 codes were used for all reviews. All analyses were performed using Atlas.ti software.

In this study, we focus on the themes and global domains found in reviews by patients who have been with their physician for at least 1 year. We describe general characteristics of Internet reviews by long-term patients and compare their comments to other reviews of primary care physicians.

Figure 1. Years with primary care physician.
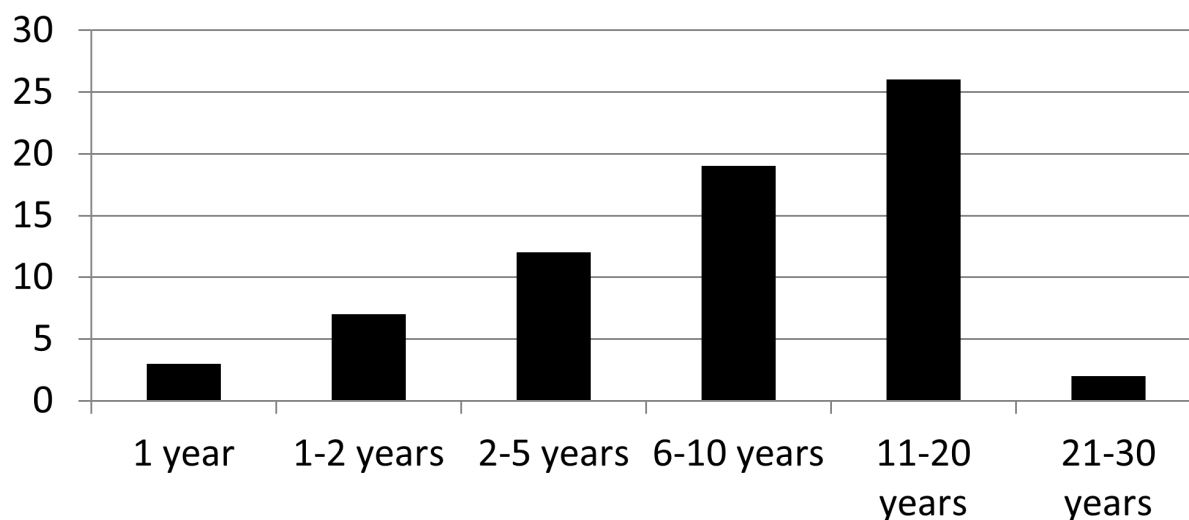
Results

Of the 712 Internet reviews of primary care physicians, a total of 93 reviews (13.1%) were from patients that self-identified as having a long-term relationship with their physician, eleven reviews (1.5%) commented on a first-time visit to a physician, and the remainder of reviews (85.4%) did not specify the amount of time with their physician. Of the reviews by long-term patients, 39% were from Yelp and 57% were from RateMDs. Long-term patients were more likely to reflect positively about their physician (86%). In contrast, only 55% of the other patients wrote positive reviews.

Analysis of long-term patient reviews also revealed six overarching domains: (1) personality traits or descriptors of the physician, (2) technical competence, (3) communication, (4) access to physician, (5) office staff/environment, and (6) coordination of care (see Figure 2). The first three domains relate directly to qualities of an individual physician while the subsequent domains reflect the physician practice and health care system (see Table 1). Overall, the reviews by long-term patients emphasized physician individual attributes. The three most prevalent themes were (1) empathy, eg, "My doctor is caring" and (2) overall excellence, eg, "Dr. X is the best", both of which fell in the domain of personality traits/descriptor, and (3) fund of knowledge, eg, "Dr. X is very knowledgeable", which is an aspect of technical competence.

Personality Traits

Most reviews by long-term patients discussed one or more physician qualities; 92% of descriptors mentioned by long-term patients were positive, and the most common themes were "amazing" and "empathetic". Other qualities frequently mentioned by long-term patients included "helpful", "professional", "calm", and "detailed". While some reviews included specific examples, many simply included a positive descriptor. Reviews by patients with either short-term or unspecified relationships with a physician also commonly

included physician descriptors, but comments were more likely to be negative (18% versus 8%). Negative descriptors included "antagonistic", "rushed", and "condescending".

Technical Competence

Physician competence was highlighted in 41% of reviews by long-term patients. These reviews discussed knowledge or clinical decision making of the physician. The patients included anecdotes describing accurate and prompt diagnosis. One reviewer remarked, "She detected my medical problem when others had missed it". Virtually all of the comments in this domain (92%) were positive.

Communication

Communication skills of the physician during a clinical encounter were described in 34% of the comments by long-term patients and 22% of all other reviews. Comments about this domain focused on physician listening skills, eg, one review stated, "[the physician] always listens to what I have to say". Other reviews referenced the ability of a physician to explain a diagnosis or new medication. One patient who remarked, "[He] explains the meds that he prescribes, he listens to and answers my questions". Notably, regardless of the length of the relationship, comments about communication were favorable (94% for long-term patient and others).

Access to Physician

We defined access as the ability to make an appointment, contact a physician, or be seen in a timely manner during a clinic visit. Descriptions of wait time and experience making an appointment were included in this domain. Many comments focused on this domain, and they were more varied than comments regarding individual attributes. While some positive comments described physicians as "accessible", negative comments about difficulty making an appointment or excessive wait times at the office were noted, even among patients with an established relationship with their physician. Reviews by long-term patients were generally favorable about experiences making an appointment

but unfavorable regarding time waiting for a scheduled appointment. One patient complained, "[I] waited almost 2 hours even though I had an appointment!" In fact, wait time was the only theme where negative comments outweighed positive comments in long-term patient reviews.

Office Staff/Environment

This domain includes all aspects of the medical visit apart from the face-to-face patient-physician interaction. Comments often referenced personality traits and helpfulness of the office staff. As an example, one patient wrote, "Staff is great—friendly and quick to respond". In this domain, there were clear differences between long-term patient reviews and other reviews, as long-term patients tended to comment favorably on nonphysician office staff (72%), while non–long-term patients complained about staff more often than giving them positive reviews. Moreover, long-term patients were less likely than other reviewers to include descriptions of the office environment (see Figure 3).

Coordination of Care

We incorporated referrals and any communication between a physician and patient outside of the individual office visit under the domain of coordination of care. Only 16% of comments by long-term patients referred to this aspect of their care, and as for physician attributes, virtually all reflections were positive. Patients described receiving prompt communication with their physician regarding laboratory test results and being pleased with the referrals their primary care physician arranged. For example, one long-term patient wrote, "When necessary, he refers me to other excellent doctors and specialists". In contrast, patients with short or unspecified relationships with a physician expressed dissatisfaction with coordination of care, as exemplified by one review that expressed, "He offered no guidance on referrals, sent me to a horrible GI". Table 2 shows the results for these three domains.

Table 1. Major themes in reviews of long-term patients.

Domain	Themes	Example Quote	Number of comments[a]	Positive comments (%)	Negative comments (%)
Personality traits					
	Empathy Overall excellence	Shows concern and competence as well as being kind, warm and friendly and respectful. I know he really cares He's a great doctor and a great person.	118	92	8
Technical competence					
	Knowledge Decision making	The guy knows his stuff - his diagnoses have always been decisive & spot on She has an excellent knowledge of medicine	38	92	8
Communication					
	Listening Explaining	I feel I can tell him or ask him ANYTHING, which is vital with your personal physician. Dr. X spends time listening to what's going on in my life and asking good questions about my health. He is a doctor who listens and talks to you like a person and not an object. He is also willing to answer any question you have and explain it in a way that a lay person can understand.	32	94	6

[a]For the number of comments, we included each instance the domain was referenced within a review. For some reviews, a domain could be mentioned more than once.

Figure 2. Conceptual model.
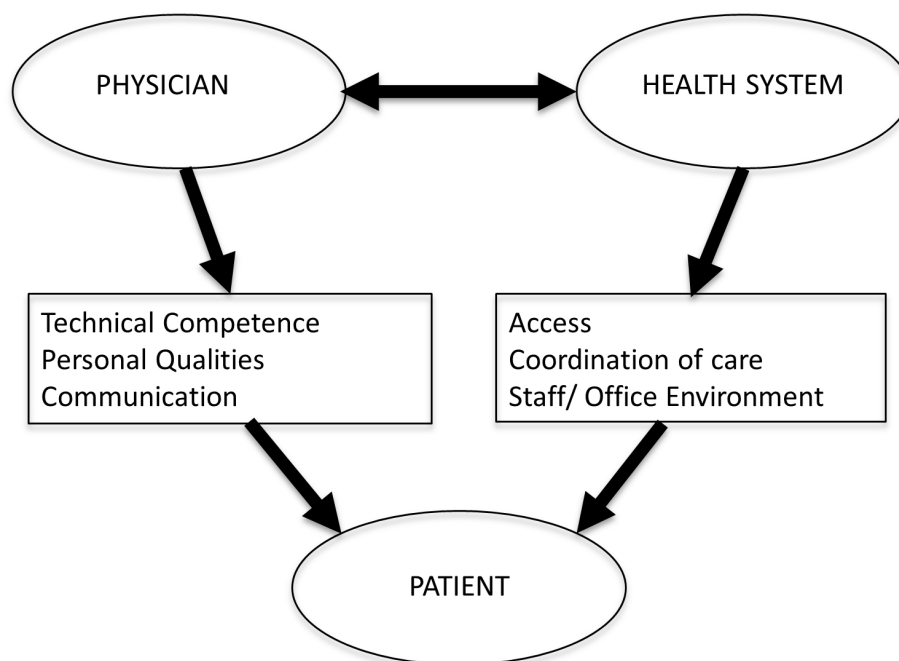
Figure 3. A comparison of long-term reviews and other reviews.
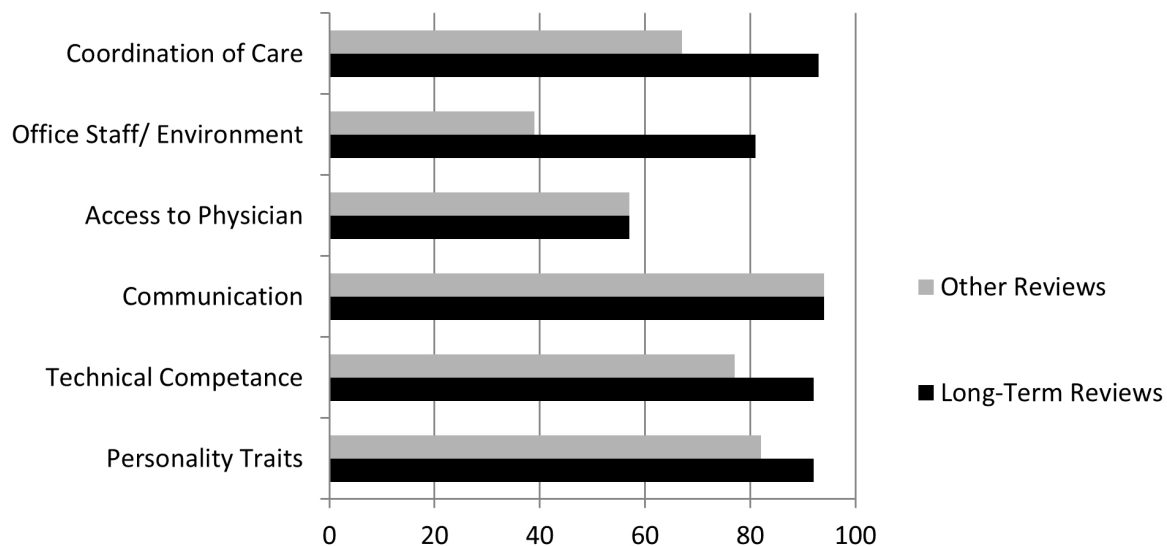
Table 2. Themes in patients' reviews.

Domain	Themes	Example Quote	Number of comments[a]	Positive comments (%)	Negative comments (%)
Access to physician					
	Making an appointment Wait time	Wait times in the office in general can be VERY long (I once waited 2 hours). Dr. X is always able to squeeze me in last minute when I am feeling sick. Night or day, he is available.	37	57	43
Staff/ office environment					
	Staff Office environment	Her office staff is great, always getting me in for an appointment after they realize what a huge worrier I am. Staff is great—friendly and quick to respond His office looks like an art gallery.	26	81	19
Coordination of care					
	Follow-up Referral—communication of test results	He called me personally with my results even though they were all normal. When necessary he refers me to other excellent doctors and specialists.	15	93	7

[a]For the number of comments, we included each instance the domain was referenced within a review. For some reviews, a domain could be mentioned more than once.

Discussion

Principal Results

Achieving continuity is important for providing quality primary care, and understanding factors patients perceive as important to long-term patient-physician relationships provides insight into promoting this continuity. Internet reviews, while limited, offer a novel perspective that can add to findings from more traditional patient satisfaction assessments. Existing patient satisfaction surveys regarding perceptions of individual primary care physicians are limited by low response rates and underrepresenting patients who are younger, poorer, less well educated, and not white [18-21]. Responders to these surveys tend to express higher satisfaction than nonresponders creating bias and overestimating patient satisfaction [31]. Our study provides unique insight into the patient's view of the patient-physician relationship and aspects that foster continuity.

Our analysis shows that patients who have been with their physician for at least 1 year write positive reviews on public websites and focus on physician attributes. Comments by established patients were more positive than other reviews, both regarding physician characteristics and technical competency. It is not surprising that patients who have been with a physician for at least 1 year write positive comments on Internet rating sites. This is consistent with previous research demonstrating an association between patient satisfaction and continuity of care [32-35]. A patient that is satisfied with encounters is more likely to return and see a given physician. Moreover, the sustained relationship likely enhances satisfaction by promoting trust and an interpersonal connection.

Personal characteristics were included in most reviews by long-term patients with positive descriptions of their physician. The positive comments about physicians' individual characteristics are consistent with other sources for evaluating patient satisfaction. This shows not only that the importance of an interpersonal connection for establishing and maintaining continuity [35-37], but also that Internet reviews reflect some similar patient values to traditional methods for measuring patient satisfaction.

The most common themes of empathy, overall excellence, and knowledge reflect aspects of medical care that promote continuity for these patients. Prior studies of patient perceptions of primary care physicians have also demonstrated the value of these factors. Empathy and patient-centered care have been associated with patient satisfaction and improved clinical outcomes [38-41]. Thus, the fact that Internet reviews also capture these factors suggest that they merit further study. Patient satisfaction is also influenced by perceived technical skill of a physician [37,42].

Factors beyond the face-to-face physician interaction also surfaced in Internet reviews. Specifically, long-term patients commented favorably on staff and office environment. While not directly influencing medical decision making, the office environment and staff may impact a patient's impression during a clinical visit. Moreover, office staff are a part of the medical

XSL•FO
RenderX

team that can facilitate or impede appropriate care. The relationship between negative perceptions of staff and patient continuity and follow-up should be specifically addressed. It is notable that long-term patients commented about staff less frequently, and it is possible that the influence of nonphysician factors wanes with duration of patient-physician relationships.

In addition, access was the most commonly included nonphysician factor in reviews by long-term patients. Previous research demonstrates that being seen within a day and having a short wait time correlates with improved patient satisfaction [35,43]. The Internet reviews show a similar emphasis, highlighting that the ability to make an appointment and be seen in a timely manner are important to patients. Time waiting in the waiting room for a given appointment was the only factor that caused dissatisfaction, regardless of the number of visits to a physician. The analysis of reviews from established patients indicate that patients are willing to tolerated suboptimal waiting times for physicians in whom they have trust and confidence. This is exemplified by one reviewer who stated, "This results in us having to wait for our appointment to be taken, but once taken, we know that she'll do a good job of helping us".

Limitations

Despite this being the first study, to our knowledge, to use public Internet-derived data to gain insight into factors associated with long-term patient-physician relationships in primary care, our findings are consistent with studies examining patient perspectives online in the context of specific health conditions like diabetes [44].

We acknowledge that our study has several limitations. First, as with all analyses using nonstandardized data, we cannot comment on the broader prevalence of the themes we uncovered in our sample. Second, as with all patient satisfaction studies, the self-selection of patients writing reviews on public websites introduces bias and may limit the generalizability of our findings. Of note, a different subset of patients are likely to complete Internet reviews than those that complete traditional patient satisfaction surveys [19,20]. Therefore, our findings may capture a novel patient perspective. Third, the majority of patients writing Internet reviews did not report the length of time with their physician. Thus, we were unlikely to have

captured all patients that were truly with their physician for longer than 1 year [32].

Despite these limitations, our findings contribute to existing knowledge regarding the patient perspective of primary care. In particular, our data show the factors important in establishing and maintaining a relationship with a physician over time.

Conclusions

Our research also adds to the data regarding public websites that enable patients to review individual physicians. Patient use of the Internet regarding health care has dramatically increased with 80% of American Internet users looking online for health information and 16% viewing reviews of health care providers [45]. The use of such websites has generated controversy both in the media and in the medical literature [15,46,47]. Research regarding the content of these websites has just begun to emerge [48-50]. Our results suggest that concerns about Internet rating affecting one's professional reputation may be overstated, as the majority of patient reviews were positive.

Website reviews of physicians are a reality and could serve as an important tool for patients as well as health care providers. Our analysis suggests ways that websites could be restructured to provide more easily accessible and reliable data. For example, differences clearly exist between individuals who have had a few visits to a physician and those with a well-established relationship with their provider. This suggests that length of time with a physician should be specified when patients write reviews. In addition, our analysis suggests that common themes emerge in reviews. Standardization of websites to direct content of reviews may also make sites more helpful to guide consumers and to guide changes in primary care practices. For physicians, reviews can provide insight into behaviors and attitudes that keep their patients engaged with care over time and also provide needed information about aspects of the visit beyond the physician-patient encounter. Factors such as staff complaints and ease of appointment-making may not be apparent to physicians but could be improved if patients' concerns were known. Further research is needed to track development of these websites, to validate structuring of website reviews, and to study how reviews impact physician practice and patient choice of their physician.

Conflicts of Interest

None declared.

References

1. Institute of Medicine. In: Kohn L, Corrigan J, Donaldson M, editors. To Err is Human: Building A Safer Health System. Washington, DC: National Academy Press; 2000.
2. Pereira AG, Pearson SD. Patient attitudes toward continuity of care. Arch Intern Med 2003 Apr 28;163(8):909-912. [doi: 10.1001/archinte.163.8.909] [Medline: 12719199]

XSL·FO

RenderX

3. Mainous AG, Goodwin MA, Stange KC. Patient-physician shared experiences and value patients place on continuity of care. Ann Fam Med 2004;2(5):452-454 [FREE Full text] [Medline: 15506580]

4. Stokes T, Tarrant C, Mainous AG, Schers H, Freeman G, Baker R. Continuity of care: is the personal doctor still important? A survey of general practitioners and family physicians in England and Wales, the United States, and The Netherlands. Ann Fam Med 2005;3(4):353-359 [FREE Full text] [doi: 10.1370/afm.351] [Medline: 16046569]

5. Gill JM, Mainous AG, Nsereko M. The effect of continuity of care on emergency department use. Arch Fam Med 2000 Apr;9(4):333-338. [Medline: 10776361]

6. O'Malley AS, Mandelblatt J, Gold K, Cagney KA, Kerner J. Continuity of care and the use of breast and cervical cancer screening services in a multiethnic community. Arch Intern Med 1997 Jul 14;157(13):1462-1470. [Medline: 9224225]

7. Parchman ML, Pugh JA, Noël PH, Larme AC. Continuity of care, self-management behaviors, and glucose control in patients with type 2 diabetes. Med Care 2002 Feb;40(2):137-144. [Medline: 11802086]

8. Saultz JW. Defining and measuring interpersonal continuity of care. Ann Fam Med 2003;1(3):134-143 [FREE Full text] [Medline: 15043374]

9. Pandhi N, Schumacher J, Flynn KE, Smith M. Patients' perceptions of safety if interpersonal continuity of care were to be disrupted. Health Expect 2008 Dec;11(4):400-408 [FREE Full text] [doi: 10.1111/j.1369-7625.2008.00503.x] [Medline: 19076668]

10. Cheng SH, Hou YF, Chen CC. Does continuity of care matter in a health care system that lacks referral arrangements? Health Policy Plan 2011 Mar;26(2):157-162 [FREE Full text] [doi: 10.1093/heapol/czq035] [Medline: 20699348]

11. Beal A, Hernandez S. Patient reports of the quality of care in community health centers: the importance of having a regular provider. J Health Care Poor Underserved 2010 May;21(2):591-605. [doi: 10.1353/hpu.0.0305] [Medline: 20453359]

12. Rosenthal TC. The medical home: growing evidence to support a new approach to primary care. J Am Board Fam Med 2008;21(5):427-440 [FREE Full text] [doi: 10.3122/jabfm.2008.05.070287] [Medline: 18772297]

13. Grumbach K, Bodenheimer T. A primary care home for Americans: putting the house in order. JAMA 2002 Aug 21;288(7):889-893. [Medline: 12186609]

14. Stange KC, Nutting PA, Miller WL, Jaén CR, Crabtree BF, Flocke SA, et al. Defining and measuring the patient-centered medical home. J Gen Intern Med 2010 Jun;25(6):601-612 [FREE Full text] [doi: 10.1007/s11606-010-1291-3] [Medline: 20467909]

15. Hodgkin PK. Web based patient feedback. BMJ 2009 Apr 07;338(apr07 1):b1377-b1377. [doi: 10.1136/bmj.b1377]

16. Jain S. Googling ourselves--what physicians can learn from online rating sites. N Engl J Med 2010 Jan 7;362(1):6-7. [doi: 10.1056/NEJMp0903473] [Medline: 20054044]

17. Mostaghimi A, Crotty BH, Landon BE. The availability and nature of physician information on the internet. J Gen Intern Med 2010 Nov;25(11):1152-1156 [FREE Full text] [doi: 10.1007/s11606-010-1425-7] [Medline: 20544300]

18. Baker R. Development of a questionnaire to assess patients' satisfaction with consultations in general practice. Br J Gen Pract 1990 Dec;40(341):487-490 [FREE Full text] [Medline: 2282225]

19. Safran DG, Karp M, Coltin K, Chang H, Li A, Ogren J, et al. Measuring patients' experiences with individual primary care physicians. Results of a statewide demonstration project. J Gen Intern Med 2006 Jan;21(1):13-21 [FREE Full text] [doi: 10.1111/j.1525-1497.2005.00311.x] [Medline: 16423118]

20. Safran DG, Kosinski M, Tarlov AR, Rogers WH, Taira DH, Lieberman N, et al. The Primary Care Assessment Survey: tests of data quality and measurement performance. Med Care 1998 May;36(5):728-739. [Medline: 9596063]

21. Taira DA, Safran DG, Seto TB, Rogers WH, Inui TS, Montgomery J, et al. Do patient assessments of primary care differ by patient ethnicity? Health Serv Res 2001 Dec;36(6 Pt 1):1059-1071 [FREE Full text] [Medline: 11775667]

22. López A, Detz A, Ratanawongsa N, Sarkar U. What patients say about their doctors online: a qualitative content analysis. Journal General Internal Medicine 2013:--- (forthcoming).

23. Donahue KE, Ashkin E, Pathman DE. Length of patient-physician relationship and patients' satisfaction and preventive service use in the rural south: a cross-sectional telephone study. BMC Fam Pract 2005 Oct 4;6:40 [FREE Full text] [doi: 10.1186/1471-2296-6-40] [Medline: 16202146]

24. Saultz JW. Defining and measuring interpersonal continuity of care. Ann Fam Med 2003;1(3):134-143 [FREE Full text] [Medline: 15043374]

25. Weiss LJ, Blustein J. Faithful patients: the effect of long-term physician-patient relationships on the costs and use of health care by older Americans. Am J Public Health 1996 Dec;86(12):1742-1747. [Medline: 9003131]

26. Rogers J, Curtis P. The concept and measurement of continuity in primary care. Am J Public Health 1980 Feb;70(2):122-127. [Medline: 7352605]

27. Borzekowski DL, Schenk S, Wilson JL, Peebles R. e-Ana and e-Mia: A content analysis of pro-eating disorder Web sites. Am J Public Health 2010 Aug;100(8):1526-1534. [doi: 10.2105/AJPH.2009.172700] [Medline: 20558807]

28. Jenssen BP, Klein JD, Salazar LF, Daluga NA, DiClemente RJ. Exposure to tobacco on the internet: content analysis of adolescents' internet use. Pediatrics 2009 Aug;124(2):e180-e186 [FREE Full text] [doi: 10.1542/peds.2008-3838] [Medline: 19620193]

XSL•FO

RenderX

29. Sans-Corrales M, Pujol-Ribera E, Gené-Badia J, Pasarín-Rua MI, Iglesias-Pérez B, Casajuana-Brunet J. Family medicine attributes related to satisfaction, health and costs. Fam Pract 2006 Jun;23(3):308-316 [FREE Full text] [doi: 10.1093/fampra/cmi112] [Medline: 16461452]

30. Beck RS, Daughtridge R, Sloane PD. Physician-patient communication in the primary care office: a systematic review. J Am Board Fam Pract 2002;15(1):25-38 [FREE Full text] [Medline: 11841136]

31. Lasek RJ, Barkley W, Harper DL, Rosenthal GE. An evaluation of the impact of nonresponse bias on patient satisfaction surveys. Med Care 1997 Jun;35(6):646-652. [Medline: 9191708]

32. Fan VS, Burman M, McDonell MB, Fihn SD. Continuity of care and other determinants of patient satisfaction with primary care. J Gen Intern Med 2005 Mar;20(3):226-233 [FREE Full text] [doi: 10.1111/j.1525-1497.2005.40135.x] [Medline: 15836525]

33. Hjortdahl P, Laerum E. Continuity of care in general practice: effect on patient satisfaction. BMJ 1992 May 16;304(6837):1287-1290 [FREE Full text] [Medline: 1606434]

34. Nutting PA, Goodwin MA, Flocke SA, Zyzanski SJ, Stange KC. Continuity of primary care: to whom does it matter and when? Ann Fam Med 2003;1(3):149-155 [FREE Full text] [Medline: 15043376]

35. Sans-Corrales M, Pujol-Ribera E, Gené-Badia J, Pasarín-Rua MI, Iglesias-Pérez B, Casajuana-Brunet J. Family medicine attributes related to satisfaction, health and costs. Fam Pract 2006 Jun;23(3):308-316 [FREE Full text] [doi: 10.1093/fampra/cmi112] [Medline: 16461452]

36. Platonova EA, Kennedy KN, Shewchuk RM. Understanding patient satisfaction, trust, and loyalty to primary care physicians. Med Care Res Rev 2008 Dec;65(6):696-712. [doi: 10.1177/1077558708322863] [Medline: 18723450]

37. Wensing M, Jung HP, Mainz J, Olesen F, Grol R. A systematic review of the literature on patient priorities for general practice care. Part 1: Description of the research domain. Soc Sci Med 1998 Nov;47(10):1573-1588. [Medline: 9823053]

38. Di Blasi Z, Harkness E, Ernst E, Georgiou A, Kleijnen J. Influence of context effects on health outcomes: a systematic review. Lancet 2001 Mar 10;357(9258):757-762. [Medline: 11253970]

39. Tarrant C, Stokes T, Baker R. Factors associated with patients' trust in their general practitioner: a cross-sectional survey. Br J Gen Pract 2003 Oct;53(495):798-800 [FREE Full text] [Medline: 14601357]

40. Little P, Everitt H, Williamson I, Warner G, Moore M, Gould C, et al. Observational study of effect of patient centredness and positive approach on outcomes of general practice consultations. BMJ 2001 Oct 20;323(7318):908-911 [FREE Full text] [Medline: 11668137]

41. Rakel DP, Hoeft TJ, Barrett BP, Chewning BA, Craig BM, Niu M. Practitioner empathy and the duration of the common cold. Fam Med 2009;41(7):494-501 [FREE Full text] [Medline: 19582635]

42. Tung YC, Chang GM. Patient satisfaction with and recommendation of a primary care provider: associations of perceived quality and patient education. Int J Qual Health Care 2009 Jun;21(3):206-213 [FREE Full text] [doi: 10.1093/intqhc/mzp006] [Medline: 19258342]

43. Anderson RT, Camacho FT, Balkrishnan R. Willing to wait?: the influence of patient wait time on satisfaction with primary care. BMC Health Serv Res 2007;7:31 [FREE Full text] [doi: 10.1186/1472-6963-7-31] [Medline: 17328807]

44. Hewitt-Taylor J, Bond CS. What e-patients want from the doctor-patient relationship: content analysis of posts on discussion boards. J Med Internet Res 2012;14(6):e155 [FREE Full text] [doi: 10.2196/jmir.2068] [Medline: 23137788]

45. Fox S. The Social Life of Health Information.: The Pew Research Center; 2011. URL: http://www.pewinternet.org/Reports/2011/Social-Life-of-Health-Info.aspx [accessed 2013-06-06] [WebCite Cache ID 6HAilpdZX]

46. Bacon N. Will doctor rating sites improve standards of care? Yes. BMJ 2009 Mar 17;338(mar17 1):b1030-b1030. [doi: 10.1136/bmj.b1030]

47. Jain S. Googling ourselves--what physicians can learn from online rating sites. N Engl J Med 2010 Jan 7;362(1):6-7. [doi: 10.1056/NEJMp0903473] [Medline: 20054044]

48. Mostaghimi A, Crotty BH, Landon BE. The availability and nature of physician information on the internet. J Gen Intern Med 2010 Nov;25(11):1152-1156 [FREE Full text] [doi: 10.1007/s11606-010-1425-7] [Medline: 20544300]

49. Reimann S, Strech D. The representation of patient experience and satisfaction in physician rating sites. A criteria-based analysis of English- and German-language sites. BMC Health Serv Res 2010;10:332 [FREE Full text] [doi: 10.1186/1472-6963-10-332] [Medline: 21138579]

50. Lagu T, Hannon NS, Rothberg MB, Lindenauer PK. Patients' evaluations of health care providers in the era of social networking: an analysis of physician-rating websites. J Gen Intern Med 2010 Sep;25(9):942-946 [FREE Full text] [doi: 10.1007/s11606-010-1383-0] [Medline: 20464523]

XSL•FO

RenderX

Original Paper

Internet Use Frequency and Patient-Centered Care: Measuring Patient Preferences for Participation Using the Health Information Wants Questionnaire

Bo Xie[1], PhD; Mo Wang[2], PhD; Robert Feldman[3], PhD; Le Zhou[2], MS

[1]School of Nursing & School of Information, University of Texas at Austin, Austin, TX, United States

[2]Department of Management, University of Florida, Gainesville, FL, United States

[3]Department of Behavioral and Community Health, University of Maryland, College Park, MD, United States

Corresponding Author:
Bo Xie, PhD
School of Nursing & School of Information
University of Texas at Austin
1710 Red River Street
Austin, TX, 78712
United States
Phone: 1 512 232 5788
Fax: 1 512 475 9179
Email: boxie@utexas.edu

Abstract

Background: The Internet is bringing fundamental changes to medical practice through improved access to health information and participation in decision making. However, patient preferences for participation in health care vary greatly. Promoting patient-centered health care requires an understanding of the relationship between Internet use and a broader range of preferences for participation than previously measured.

Objective: To explore (1) whether there is a significant relationship between Internet use frequency and patients' overall preferences for obtaining health information and decision-making autonomy, and (2) whether the relationships between Internet use frequency and information and decision-making preferences differ with respect to different aspects of health conditions.

Methods: The Health Information Wants Questionnaire (HIWQ) was administered to gather data about patients' preferences for the (1) amount of information desired about different aspects of a health condition, and (2) level of decision-making autonomy desired across those same aspects.

Results: The study sample included 438 individuals: 226 undergraduates (mean age 20; SD 2.15) and 212 community-dwelling older adults (mean age 72; SD 9.00). A significant difference was found between the younger and older age groups' Internet use frequencies, with the younger age group having significantly more frequent Internet use than the older age group (younger age group mean 5.98, SD 0.33; older age group mean 3.50, SD 2.00; t_{436}=17.42, P<.01). Internet use frequency was positively related to the overall preference rating (γ=.15, P<.05), suggesting that frequent Internet users preferred significantly more information and decision making than infrequent Internet users. The relationships between Internet use frequency and different types of preferences varied: compared with infrequent Internet users, frequent Internet users preferred *more* information but *less* decision making for diagnosis (γ=.57, P<.01); *more* information and *more* decision-making autonomy for laboratory test (γ=.15, P<.05), complementary and alternative medicine (γ=.32, P<.01), and self-care (γ=.15, P<.05); and *less* information but *more* decision-making autonomy for the psychosocial (γ=-.51, P<.01) and health care provider (γ=-.27, P<.05) aspects. No significant difference was found between frequent and infrequent Internet users in their preferences for treatment information and decision making.

Conclusions: Internet use frequency has a positive relationship with the overall preferences for obtaining health information and decision-making autonomy, but its relationship with different types of preferences varies. These findings have important implications for medical practice.

XSL·FO
RenderX

KEYWORDS

patient-centered care; patient preference; shared decision-making; patient participation; Internet

Introduction

Patient participation in health care decision-making has both legal and ethical grounds [1]. It is increasingly recognized as a cornerstone of patient-centered health care that can improve health care quality and outcomes [2-4] and reduce utilization of health care resources [5] and costs [6,7]. The recent shift from a paternalistic to a shared or informed model of health care decision-making [8-14] has drawn much attention to patient participation, although there is little consensus regarding exactly what patient participation entails [15]. Patient preferences or desire for the amount of information about different aspects of a health condition and level of decision-making autonomy across different aspects of a health condition are commonly used as two major indicators of patient participation, eg, [11,12,16]. However, these two types of preferences are often measured differently across studies [17], making it difficult to compare reported findings.

The Breadth of Patient Preferences for Participation

The instruments commonly used for measuring patient preferences, established well before the prevalence of Internet use in contemporary health care, focus on a limited range of types of health information and an even more limited range of types of decision making. All of the commonly used instruments measuring preferences for obtaining health information include measures of preferences for obtaining information about treatment and diagnosis [16,18-24]. Several also include measures of preferences for obtaining information about laboratory testing/medical examination [16,18,21-24] and physical/self-care [20,22-24], but only two include measures of preference for obtaining psychosocial information [20,22]. Meanwhile, the instruments commonly used for measuring preferences for decision-making autonomy all measure primarily or even exclusively preference for participation in (standard) *treatment* decision making [16,18,19,23,25]. Other types of decision making, such as decisions regarding what medical facility to use or whether to seek complementary or alternative treatments, are understudied or even completely missing from these widely used instruments.

Currently, there is no known validated instrument measuring preferences for obtaining online health information or decision-making autonomy based on the information obtained online. However, Internet studies have found a broader range of preferences for obtaining health information and decision making autonomy than found in earlier studies [26-32]. For instance, while information about diagnosis and treatment still comprises the main types of health information that older adults seek online, several other types of health information (eg, information about nutrition, exercise, and body weight; health care providers; and alternative treatments) are also commonly sought online by older Internet users [33]. Also, using information obtained online, individuals are making a wide range of decisions regarding, eg, treatment, health care facilities and providers, how to interact with physicians (eg, what

questions to ask and how to ask during an office visit), how to cope with a condition, and how to think about healthy eating, exercise, or stress management [34-38]. Some Internet studies have even revealed Internet users making decisions regarding diagnosis based on the information they obtained online [28,39].

This new broader coverage of the types of health information and decision making has helped to reveal interesting phenomena previously understudied or ignored. It also calls for a more systematic examination of the relationship between Internet use and a broad range of information and decision-making preferences.

Measuring Preferences for Participation: The Health Information Wants Questionnaire

Derived from a grounded theory study, our health information wants (HIW) framework encompasses a broad range of types of information and decision making and presents each type of information as corresponding to one type of decision making [40]. Building on and further testing the HIW framework, we developed the Health Information Wants Questionnaire (HIWQ) through a multistage process over the course of 2 years [17,41]. The HIWQ differs from prior instruments in at least three important ways. First, it measures preferences for seven types of health information and decision making—information and decision making about diagnosis, treatment, laboratory testing, self-care, complementary and alternative medicine (CAM), psychosocial aspect, and health care providers. Second, the items on the information dimension parallel those on the decision-making dimension (ie, each item on the Information Scale has a corresponding, parallel item on the Decision Making Scale), making it possible to more directly compare preferences for participation in different types of information seeking and decision making. Finally, the HIWQ has a built-in consideration for exploring potential impacts of Internet use frequency on preferences for obtaining health information and decision-making autonomy. Detailed descriptions of the development process of the HIWQ, including our rationale for focusing on these seven types of information/decision making and the selection and development of the specific items within each type, are reported elsewhere [17,42].

In this paper, we report findings from the first large sample study using the HIWQ, focusing specifically on the relationship between *Internet use frequency* and preferences among undergraduate students and older adults. We selected these two particular age groups mainly because of the sharp contrast between their Internet use frequencies: the younger age group typically has the highest level of Internet use frequency, whereas the older adult age group has the lowest [43]. Findings from the same large sample study focusing on the relationship between age and each type of preference are reported elsewhere [42].

Research Questions

Previous research has suggested that factors such as age, gender, education, culture, the role of being a patient, severity of health condition, and personality are related to patients' preferences

for participation in their own health care [16,19,21,44-50]. Given the accumulating amount of evidence in the literature suggesting connections between Internet use and patient participation, we asked the following primary research question (RQ):

RQ1: Is there a significant relationship between Internet use frequency and the overall preferences for obtaining health information and decision-making autonomy?

Previous research has indicated that preferences for participation are highly variable [51-55]. However, to date there is little knowledge about how different Internet users may have different preferences for participation. Recognizing this gap in the literature, we asked another primary RQ:

RQ2: Does the relationship between Internet use frequency and information and decision-making preferences differ with respect to seven different aspects of health conditions—diagnosis, treatment, laboratory testing, self-care, CAM, psychosocial aspect, and health care providers?

Methods

Participants

A convenience sample of 438 individuals participated in this study. Participants included 226 undergraduate students majoring in a variety of disciplines at a large state university and 212 older adults recruited from senior-oriented computer classes held at public libraries and senior centers. Participants were recruited through flyers posted in building hallways and message boards, advertisements in local newspapers, and word of mouth. Demographic characteristics of the participants are reported in Table 1 (following the Health and Retirement Study [56], we coded eight conditions—high blood pressure, diabetes, cancer, lung disease, heart disease, stroke, psychiatric problems, and arthritis—as "major" health conditions and all other conditions as "minor" health conditions).

Materials

The data reported here were obtained using the 21-item HIWQ. This 21-item instrument is a psychometrically improved version (in terms of both reliability and construct validity) of our original 40-item HIWQ [41]. In addition, it significantly shortens the time required by participants to complete it. This self-administered instrument includes two main scales: the Information Preference Scale and the Decision Making Preference Scale. Each of these scales contains seven subscales with parallel items in the following information and decision-making categories: diagnosis (items 1-4), treatment (items 5-7), laboratory testing (items 8-10), self-care (items 11-13), CAM (items 14-16), psychosocial aspect (items 17-19), and health care providers (items 20-21) in the information and decision-making subscales (Multimedia Appendix 1).

On the Information Preference Scale, participants indicated their preferences for each type of information (eg, How much information would you like to have about how severe a health condition is) on a 5-point Likert-type scale, in which response choices ranged from 1 (None) to 5 (All). On the Decision Making Preference Scale, participants also indicated their preferences for each type of health decision making on a 5-point

Likert-type scale (eg, Who do you think should make the decision regarding how severe a health condition is). Adapted from Ende et al [16], response choices were the doctor alone (1), mostly the doctor (2), the doctor and myself equally (3), mostly myself (4), and myself alone (5).

In addition to the 21 parallel items on the Information and Decision Making Scales, the HIWQ also included items measuring age (younger vs older), gender (male vs female), general health status, health condition (major vs minor), whether the condition was current or past, how long the condition lasted, severity of the condition, how knowledgeable the participant was about the condition, marital status, education level, ethnicity, income level, and Big Five personality (extraversion, agreeableness, conscientiousness, neuroticism, and openness). As summarized in several review articles [51-55], these variables were found to be related to preferences for obtaining health information and decision-making autonomy. These variables were therefore used as control variables in all relevant analyses reported here.

Before completing the Information and Decision Making Scales, participants were asked to first think about a specific health condition that they had in the past or currently had and to continue thinking about this specific health condition while filling out the rest of the questionnaire.

Procedure

Completion of the instrument took place in a quiet university classroom or office for the undergraduate participants and in a quiet meeting room in a public library or senior center for the older participants. Prior to data collection, all participants completed the informed consent form, approved by the Institutional Review Board of the authors' university. Participants were instructed to complete the instrument independently, using paper and pen. On average, it took approximately 15-25 minutes for an undergraduate participant and 30-45 minutes for an older adult to complete the instrument. Data collection took place from May to December 2010.

Data Analysis

Data in the current study had a nested structure in which each participant rated items in two dimensions (ie, information preference and decision-making preference). The subscale and overall dimension scores were first calculated as means across relevant items. Following the strategy used by Ende et al [16], these original scores were then rescaled to have a midpoint of 50 and ranges from 0 (corresponding to least desire for information seeking or decision making) to 100 (corresponding to strongest desire for information seeking or decision making). The rescaling was done by linearly transforming the original score, ie, rescaled score=(raw score-1)*25. This rescaling strategy allowed us to compare the scores of the information and decision-making dimensions. Internet use frequency was between-subject level (ie, Level 2) predictor whereas dimension of preference ratings was a within-subject level predictor (ie, Level 1). Preference rating was the outcome variable. Since Internet use frequency is a continuous variable, repeated-measure ANOVA is not appropriate for testing its interaction effect with rating dimension on preference ratings.

Therefore, we used the multilevel modeling technique [57] to estimate the interaction effect of Internet use frequency and rating dimension on preference ratings. Dimension of preference was coded as a dummy variable with decision-making preference = "0" and information preference = "1", which had a random effect on preference ratings. Internet use frequency was treated as Level-2 predictor, which had effects on the random intercept of preference ratings and on the random slope of the dimension-rating relationship. In addition, we controlled for the main effects of age group, gender, general health status, whether had health condition in the past or current, how long had the condition, severity of the condition, knowledge of the condition, marital status, education, ethnicity, income, and Big Five personalities on preference ratings in the model. (Gender was coded as 1=male and 0=female. Health condition was coded as 1=major and 0=minor. Condition time was coded as 1=current and 0=past. Marital status was coded by dummy coding scheme, with married as the referent group. Ethnicity was coded by dummy coding scheme, with white as the referent group.)

Table 1. Demographic characteristics of study participants.

Variable		Young n=226	Older n=212	Total n=438
Age				
	Minimum	18	50	18
	Maximum	32	100	100
	Mean	20.31	71.92	44.16
	SD	2.15	9.00	26.52
Gender, n (%)				
	Female	165 (73.0)	139 (65.6)	304 (69.4)
	Male	61 (27.0)	73 (34.4)	134 (30.6)
Marriage status, n (%)				
	Married	2 (.9)	72 (34.0)	74 (16.9)
	Single	217 (96.0)	30 (14.1)	247 (56.4)
	Separated	2 (.9)	4 (2.0)	6 (1.4)
	Divorced	1 (.4)	32 (15.1)	33 (7.5)
	Widowed	3 (1.3)	74 (34.7)	77 (17.6)
	Living as married	1 (.4)	0 (0)	1 (.2)
Highest level of education, n (%)				
	Less than high school graduate	0 (0)	9 (4.2)	9 (2.1)
	High school graduate/GED	72 (31.9)	63 (29.7)	135 (30.8)
	Vocational training	1 (.4)	13 (6.1)	14 (3.2)
	Some college/associate's degree	135 (59.7)	56 (26.4)	191 (43.6)
	Bachelor's degree	17 (7.5)	35 (16.5)	52 (11.9)
	Master's degree or other postgraduate training	1 (.4)	30 (14.2)	31 (7.1)
	Doctoral degree	0 (0)	6 (2.8)	6 (1.4)
Membership in ethnic group, n (%)				
	Asian	22 (9.7)	11 (5.2)	33 (7.5)
	African American	117 (51.8)	105 (49.5)	222 (50.7)
	Latino/Hispanic	8 (3.5)	8 (3.8)	16 (3.7)
	Native American/American Indians/Alaska Native	1 (0.4)	2 (0.9)	3 (0.7)
	Native Hawaiian/Pacific Islander	0 (0)	2 (0.9)	2 (0.5)
	White	78 (34.5)	84 (39.6)	162 (37.0)
Annual household income, n (%)				
	Less than $20,000	56 (24.8)	45 (21.2)	101 (23.1)
	$20,000-$29,999	7 (3.1)	31 (14.6)	38 (8.7)
	$30,000-$39,999	7 (3.1)	42 (19.8)	49 (11.2)
	$40,000-$49,999	7 (3.1)	30 (14.2)	37 (8.4)
	$50,000-$59,999	9 (4.0)	26 (12.3)	35 (8.0)
	$60,000-$69,999	17 (7.5)	14 (6.6)	31 (7.1)
	$70,000-$99,999	23 (10.2)	14 (6.6)	37 (8.4)
	More than $99,999	100 (44.2)	10 (4.7)	110 (25.1)
Health condition, n (%)				
	Major	36 (15.9)	134 (63.2)	170 (38.8)

Variable	Young n=226	Older n=212	Total n=438
Minor	190 (84.1)	78 (36.8)	268 (61.2)
When had the condition, n (%)			
Past	114 (50.4)	63 (29.7)	177 (40.4)
Current	112 (49.6)	149 (70.3)	261 (59.6)

Results

Psychometrics

The results suggest that the overall Information Scale, the overall Decision Making Scale, and all the subscales of these two scales were internally consistent and reliable (Cronbach alpha coefficients ranged from .95-.71 for the younger age group, and .98-.78 for the older age group); confirmatory factor analyses supported the construct validity of the HIWQ (see [42] for detailed descriptions of the reliability and construct validity of the instrument). Furthermore, the overall scores for both the Information and Decision Making Scales were significantly correlated with those for their corresponding global items (for Information, "How much information would you like to have about this condition?"; For Decision Making, "Who do you think should make the decision related to this specific health condition?"). Specifically, for young adults, the correlation was .42 ($P<.01$) for the information dimension and .34 ($P<.01$) for the decision-making dimension. For older adults, the correlation was .61 ($P<.01$) for the information dimension and .49 ($P<.01$) for the decision-making dimension. These significant correlations support the convergent validity of the HIWQ.

Internet Use Frequency

Internet use frequency was measured by the following item: How often do you use the Internet? Responses ranged from Never (1) to Everyday (6). Significant difference was found between the younger and older age groups' Internet use frequencies, with the younger age group having significantly more frequent Internet use than the older age group (younger age group mean 5.98, SD 0.33; older age group mean 3.50, SD 2.00; t_{436}=17.42, $P<.01$).

Relationship Between Internet Use Frequency and Overall Preferences

Results of multilevel modeling analysis (Tables 2 and 3) showed that, after controlling for age group, gender, general health status, health condition (major vs minor), whether the condition was current or past, how long the condition lasted, severity of the condition, how knowledgeable participants were about the condition, marital status, education level, ethnicity, income level, and Big Five personality, Internet use frequency was positively related to the overall preference rating (γ=.15, $P<.05$), suggesting that frequent Internet users preferred significantly more information and decision-making autonomy than did infrequent Internet users. Internet use frequency did not predict the random slope between rating dimension (information vs decision making) and the overall preference rating, suggesting that there was no interaction among Internet use frequency,

overall information preference, and overall decision-making preference.

Relationship Between Internet Use Frequency and Each Type of Preference

In the following analyses, the main effects of age group, gender, general health status, health condition, whether the condition was current or past, how long the condition lasted, severity of the condition, how knowledgeable participants were about the condition, marital status, education level, ethnicity, income level, and Big Five personality on preference ratings were controlled for. The results of multilevel modeling analysis for the relationship between Internet use frequency and each type of preference are also reported in Tables 2 and 3.

For the diagnosis subscale, the main effect of Internet use frequency on preference rating was not significant. However, results of multilevel modeling analysis showed that Internet use frequency was positively related to the random slope between rating dimension (information vs decision making) and preference rating (γ=.57, $P<.01$), suggesting an interaction effect of Internet use frequency on this rating dimension. These results indicated that frequent Internet users preferred obtaining more information but less decision-making autonomy about diagnosis than did infrequent Internet users (Figure 1).

For the psychosocial subscale, the main effect of Internet use frequency on preference rating was not significant. However, results of multilevel modeling analysis showed that Internet use frequency was negatively related to the random slope between rating dimension (information vs decision making) and preference rating (γ=-.51, $P<.01$), suggesting an interaction effect of Internet use frequency on this rating dimension. These results indicated that frequent Internet users preferred obtaining less information but more decision-making autonomy about psychosocial aspects than did infrequent Internet users (Figure 2).

For the health care provider subscale, the main effect of Internet use frequency on preference rating was not significant. However, results of multilevel modeling analysis showed that Internet use frequency was negatively related to the random slope between rating dimension (information vs decision making) and preference rating (γ=-.27, $P<.05$), suggesting an interaction effect of Internet use frequency on this rating dimension. These results indicated that frequent Internet users preferred obtaining less information but more decision-making autonomy about health care providers than did infrequent Internet users (Figure 3).

Results of multilevel modeling analysis showed that Internet use frequency was positively related to preference rating for the

laboratory test (γ=.15, P<.05), self-care (γ=.15, P<.05), and CAM (γ=.32, P<.01) subscales. For these subscales, Internet use frequency did not predict the random slope between rating dimension (information vs decision making) and preference rating. These results suggested that frequent Internet users would prefer obtaining more information and decision-making autonomy about laboratory testing, self-care, and CAM than infrequent Internet users would. For the treatment subscale, Internet use frequency was not significantly related to preference rating or the random slope between rating dimension and ratings.

Table 2. Multilevel modeling results – overall, diagnosis, treatment, and laboratory test (Level 2 [ie, between-person level] N=438; Level 1 [ie, within-person level] N=876; unstandardized coefficients are reported).

| Variable	Overall	Diagnosis	Treatment	Laboratory test
Random intercept (β_0)				
Intercept (γ_{00})	5.68 [a]	4.81 [a]	5.69 [a]	4.57 [a]
Age group (γ_{01})	.37	1.09 [a]	.17	1.16 [a]
Gender (γ_{02})	-.20	-.21	-.22	.08
Health condition (γ_{03})	-.15	-.29	-.29	-.20
Condition time (γ_{04})	.00	.05	.21	.05
Years of condition (γ_{05})	.01	.00	.01	.00
Severity (γ_{06})	.06	.11	.16	.11
Knowledgeable (γ_{07})	-.00	-.04	.08	.06
General health status (γ_{08})	-.05	-.06	-.16	-.19
Education (γ_{09})	-.02	.04	.03	-.02
Single vs married (γ_{010})	.06	.25	.43	.25
Separated vs married (γ_{011})	.09	.32	.73	-.32
Divorced vs married (γ_{012})	.46	.29	.73 [b]	.43
Widowed vs married (γ_{013})	.10	.22	.31	-.05
Living as married vs married (γ_{014})	.06	-.06	.83	-2.62 [a]
Asian vs white (γ_{015})	-.13	.09	-.26	.10
African American vs white (γ_{016})	-.28 [b]	-.11	-.03	-.07
Latino vs white (γ_{017})	.18	-.08	-.56	.61
Native American vs white (γ_{018})	-.91	-.15	.68	-.97 [a]
Pacific Islander vs white (γ_{019})	-.12	-.37	-1.18	.27
Income (γ_{020})	-.02	-.01	-.02	-.07 [b]
Extraversion (γ_{021})	.01	.13 [a]	.03	.04
Agreeableness (γ_{022})	-.01	.01	.01	.00
Conscientiousness (γ_{023})	.04	.03	-.02	.03
Neuroticism (γ_{024})	.02	.11 [b]	.09	.03
Openness (γ_{025})	.01	-.01	.11	-.01
Internet use frequency (γ_{026})	.15 [b]	.11	.15	.15 [b]
Residual variance ($\upsilon_I{}^2$)	.88	.69	.36 [b]	1.07
Random slope for preference dimension (β_1)				
Intercept (γ_{10})	2.77 [a]	4.95 [a]	4.53 [a]	4.98 [a]
Internet use frequency (γ_{11})	.02	.57 [a]	.07	.06
Residual variance ($\upsilon_I{}^2$)	5.71 [a]	8.58 [a]	5.15 [a]	6.91 [a]
Level 1 residual variance (σ^2)	1.22 [a]	2.38 [a]	3.38 [a]	2.98 [a]

[a]$P<.01$.

[b]$P<.05$.

XSL•FO
RenderX

Table 3. Multilevel modeling results – self-care, CAM, psychosocial, and health care provider (Level 2 [ie, between-person level] N=438; Level 1 [ie, within-person level] N=876; unstandardized coefficients are reported).

Variable	Self-care	CAM	Psychosocial	Health care provider
Random intercept (β_0)				
Intercept (γ_{00})	6.45 [b]	5.79 [b]	6.44 [b]	6.45 [b]
Age group (γ_{01})	-.71 [a]	.62	-.99 [b]	.10
Gender (γ_{02})	-.18	-.24	-.24	-.72 [b]
Health condition (γ_{03})	-.12	-.18	-.01	-.09
Condition time (γ_{04})	-.09	-.03	-.01	-.13
Years of condition (γ_{05})	.02	.02	.02 [a]	.02
Severity (γ_{06})	.18 [a]	.09	.07	.08
Knowledgeable (γ_{07})	-.12	-.02	-.02	-.04
General health status (γ_{08})	-.09	-.15	.06	.05
Education (γ_{09})	-.03	-.01	.03	-.08
Single vs married (γ_{010})	-.44	.01	-.45	-.03
Separated vs married (γ_{011})	-.57	.59	.43	-.07
Divorced vs married (γ_{012})	.19	.85 [a]	.48	.55
Widowed vs married (γ_{013})	-.40	.32	.14	-.05
Living as married vs married (γ_{014})	-1.21 [a]	2.19 [b]	-.78	-1.78 [b]
Asian vs white (γ_{015})	-.29	-.25	-.31	-.07
African American vs white (γ_{016})	-.48 [a]	-.46 [a]	-.33 [a]	-.16
Latino vs white (γ_{017})	.09	-.02	.08	.28
Native American vs white (γ_{018})	-1.63	-.91	-1.32	-.88
Pacific Islander vs white (γ_{019})	1.28	-.60	.63	-1.43
Income (γ_{020})	-.02	-.03	.02	-.01
Extraversion (γ_{021})	-.13 [a]	-.01	-.05	-.01
Agreeableness (γ_{022})	.01	.05	-.03	.07
Conscientiousness (γ_{023})	.08	.10	.02	-.02
Neuroticism (γ_{024})	-.15 [a]	.04	-.10	.09
Openness (γ_{025})	.05	.04	-.02	.04
Internet use frequency (γ_{026})	.15 [a]	.32 [b]	-.04	.13
Residual variance ($\upsilon_1{}^2$)	.97	.93	1.67	1.57
Random slope for preference dimension (β_1)				
Intercept (γ_{10})	2.18 [b]	2.21 [b]	-2.16 [b]	1.20 [b]
Internet use frequency (γ_{11})	-.04	.07	-.51 [b]	-.27 [a]
Residual variance ($\upsilon_1{}^2$)	7.34 [b]	10.56 [b]	14.42 [b]	9.19 [b]
Level 1 residual variance (σ^2)	γ	γ	2.46 [b]	4.47 [b]

[a] $P<.05$.

[b] $P<.01$.

Figure 1. Interaction between Internet use frequency and rating dimension (Information vs Decision Making) for the Diagnosis Subscale.
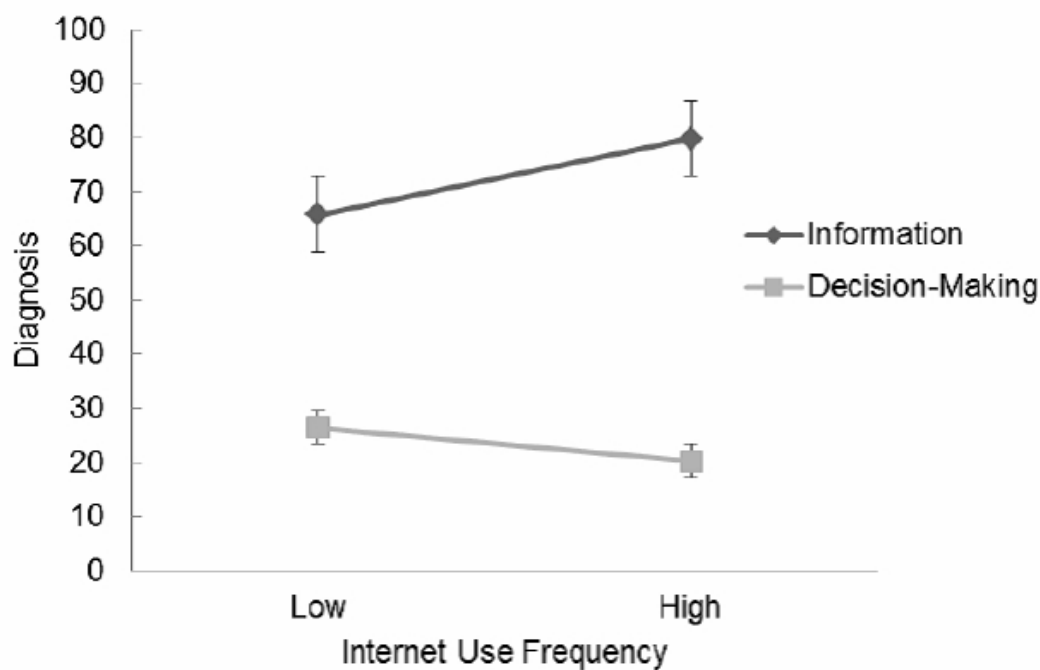
Figure 2. Interaction between Internet use frequency and rating dimension (Information vs Decision Making) for the Psychosocial Subscale.
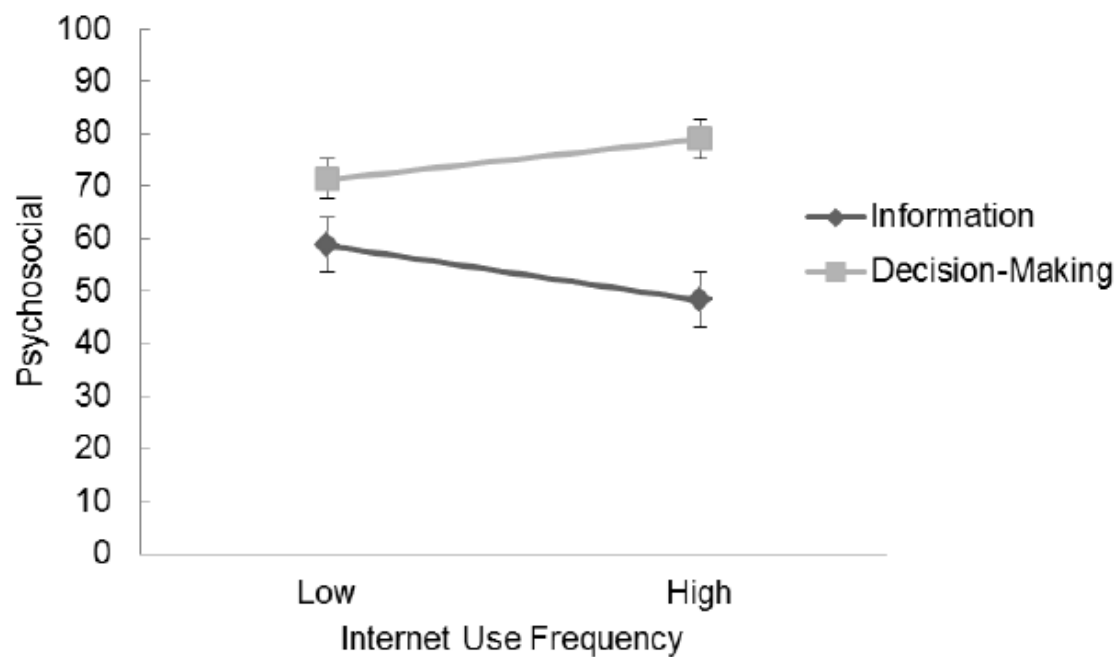
Figure 3. Interaction between Internet use frequency and rating dimension (Information vs Decision Making) for the Health Care Provider Subscale.
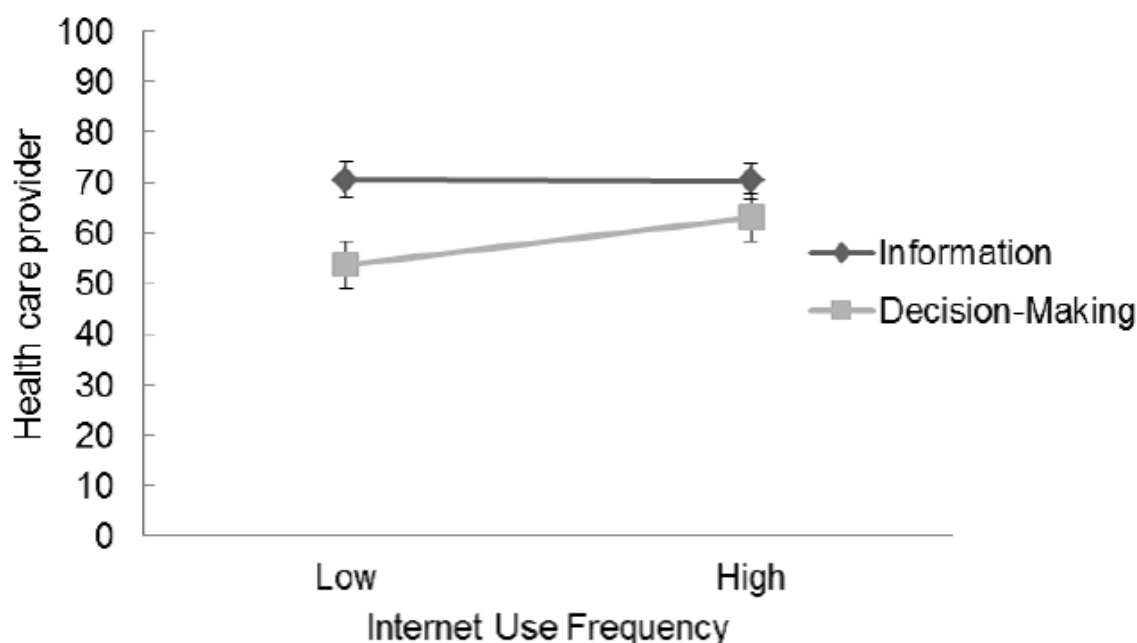
Discussion

Principal Findings

Promoting patient-centered health care requires an understanding of patient preferences for obtaining health information and decision-making autonomy. Recent developments in information and communication technologies have introduced complications to the scope and extent of patient participation [58-60]. Some argue that the Internet is bringing fundamental changes to the medical profession [58,61]. Drawing upon Paul Starr's framework of medical professionalism [62], Blumenthal [58] has argued that the Internet has enabled patients to challenge two particular core attributes contributing to the distinctive competence of medical professionals. First, the Internet creates unprecedented opportunities for the general public to access vast amounts of medical knowledge previously known only to medical professionals, thus challenging the *cognitive* attribute of the medical profession. This argument is supported by empirical studies showing a large number of health consumers obtaining health information from the Internet [63,64]. Second, by generating convenient access to information about the credentials and experiences of medical professionals [65], the Internet also enables the general public to make informed decisions about the track record of their physicians [27,66], thus challenging the *collegial* attribute of the medical profession (ie, self-monitoring and self-discipline within the profession itself) [58]. Ample empirical evidence supports this argument. For instance, through various online tools including social media sites, health consumers are actively describing, rating, and sharing their experiences of health care facilities and physicians, and on the basis of peer experience, making decisions regarding which facility or physician to go to [67-69]. In fact, this bottom-up approach has become so prevalent that medical professionals have begun to explore how to make best use of such patient-generated ratings and content [70,71].

The findings of the present study provide further empirical evidence for these arguments by revealing a positive correlation between Internet use and patient participation. Specifically, with regard to RQ1 (Is there a significant relationship between Internet use frequency and the overall preferences for obtaining health information and decision-making autonomy?), our findings show that Internet use frequency was positively related to overall preference rating, suggesting that frequent Internet users preferred significantly more information and decision-making autonomy than did infrequent Internet users. Interestingly, findings from this study (reported elsewhere) also suggest that age was not associated with overall preference rating [42]. Therefore, compared with age, Internet use frequency appears to be more strongly associated with overall preference for health information and decision-making autonomy in this study. These findings have important implications for medical practice: when medical professionals attempt to gauge how much information to provide to patients or try to decide how much they should involve patients in medical decision-making, they may be better off if they base their decisions on patients' Internet use frequency rather than age per se.

With regard to RQ2 (Does the relationship between Internet use frequency and information and decision-making preferences differ with respect to seven different aspects of health conditions, ie, diagnosis, treatment, laboratory testing, self-care, CAM, psychosocial aspect, and health care providers?), our findings suggest that the relationship between Internet use

frequency and different types of preferences varies. Specifically, compared with infrequent Internet users, frequent Internet users preferred *more* information but *less* decision-making autonomy for diagnosis, *more* information and *more* decision-making autonomy for laboratory testing, CAM, and self-care, and *less* information but *more* decision-making autonomy for the psychosocial and health care provider aspects. For treatment, we did not find a significant difference between frequent and infrequent Internet users in their information and decision-making preferences.

These findings challenge others widely reported in the literature. In particular, there seems to be a consensus that patients are interested in obtaining more information but are not as interested in participating in decision-making [25,44,72-75]. The context of such a "consensus" though, as we have explained, is the fact that previously only a very limited range of preferences was measured, while other types of preferences—that might not be perceived as important by medical professionals but nonetheless are important from the patient's perspective—were largely ignored [17,42]. Using the HIWQ, which covers a broader range of preferences than previous instruments and presents parallel items on the information and decision-making scales, we have been able to develop a more comprehensive view of patient preferences consisting of nuances previously ignored.

These nuances have important implications for medical practice, particularly given the increasing emphasis on patient-centered health care [3]. For instance, our findings suggest that Internet use frequency is positively associated with overall preference for health information and participation in decision-making, but that when overall preference is broken down into different aspects, the relationship between Internet use frequency and different types of preferences varies from one aspect to another. Thus, to encourage patient participation, medical professionals might want to consider promoting different aspects of participation to different extents to better accommodate patients' preferences. For instance, medical professionals might want to provide frequent Internet users with more information and more decision-making autonomy about laboratory testing, CAM, and self-care than they would provide to infrequent Internet users. However, medical professionals might not need to provide as much psychosocial information for frequent Internet users as for infrequent Internet users.

Previous research suggests that age is a strong predictor of patient preferences [16], with younger adults having a significantly stronger desire for both information and decision-making autonomy than their older counterparts [16,19,21,44-47]. However, our findings suggest that age was not associated with the overall preference rating or preference about treatment and CAM; furthermore, on the subscales where age was related to preference ratings (diagnosis, psychosocial aspect, health care providers, and self-care), its effect is in line with that of Internet use frequency [42]. These findings suggest that, just as when they make decisions regarding *overall* information and decision-making preference, medical professionals, when they try to decide how much of a specific type of information to provide to patients or how much to involve patients in specific types of decision-making, may want

to base their decisions on patients' Internet use frequency rather than age.

Limitations and Future Directions

This study used a convenience sample. Considering that some of the relationships tested were statistically significant, size of the current sample did allow sufficient statistical power for testing the effects of interest. Still, the results may not be representative. Caution should be taken in generalizing the findings to the general population. The sample consisted of two groups, undergraduates 18-32 years old and older adults 50-100; these groups were frequent and infrequent Internet users, respectively. Additional research should address a broader range of Internet use frequency to determine whether these results could be replicated across groups with varying levels of Internet use frequency (and it would be especially interesting to compare and contrast older adults who are *frequent* Internet users with younger adults who are *infrequent* Internet users to better understand the relationships among age, Internet use frequency, and preference for participation). Furthermore, in this study we measured the construct of "Internet use frequency", which is a subconstruct of "Internet use" that may involve broader variation than the "frequency" of use. It would be interesting in future research to further validate the findings in a population of patients seeking care whose interest in technology and actual use of it may vary more widely than the two populations (ie, older adults at a computer class and college students) examined in this study.

The HIWQ, when administered in cross-sectional studies like the present one, provides only a snapshot view of preferences. Yet, experiences of illness can span months or even years, and preferences for obtaining health information and decision-making autonomy may change over time [76-78]. In future research, it will be necessary to administer the HIWQ multiple times to assess and compare if and how patient preferences for participation might evolve over the course of their conditions. Another limitation is that some of the decision-making subscales showed lower Cronbach alpha values in the younger age group [42]. One possible reason is that the younger participants had less life experience with making important medical decisions. Therefore, the constructs and the items were less familiar to them, which might lead to lower Cronbach alphas. Future research should further investigate this issue by collecting data from other younger adult samples. Additionally, in our study, we had only one global item for the information scale and one item for the decision-making scale. Future research may use another measure with multiple items for each of these scales to provide more persuasive evidence for the instrument's convergent validity. Finally, as reviewed above, patient preferences are often used in the literature as indicators of patient participation in their own health care. However, preferences may not already be a perfect proxy for actual participation. Further research would need to confirm correlation between preference and actual participation in health information seeking and decision making.

Conclusions

Internet applications have created unprecedented opportunities for patient participation through improved access to a wide

range of health information previously difficult for the general public to obtain [33,63-65]. Patients are now better equipped with knowledge necessary to make more informed decisions about a broad range of health care-related issues [27,28,34-39,66-69]. Not surprisingly, it has been suggested that the Internet is bringing fundamental changes to the medical profession [58,61], as patients become more informed, more participatory, and consequently, more empowered [79]. Our findings, while supporting this general argument about the relationship between Internet use frequency and patient participation and empowerment, also reveal novel nuances in this relationship (eg, when patient preference is broken down into seven aspects, the relationships between Internet use frequency and type of information preference and its

corresponding decision-making preference clearly vary across those aspects).

Previous research suggests that age, gender, education, culture, the role of being a patient, severity of health condition, and personality can help explain the variance in patient preferences [16,19,21,44-50]. This study reveals a new related factor for patient preferences: Internet use frequency, which was significantly related to not only overall preference but also preferences for several types of information and decision-making autonomy. These findings may have important implications for medical practice. For example, medical professionals may want to take into account their patients' Internet use frequency when understanding if, how much, and in what ways their patients might wish to participate in their own health care.

Acknowledgments

Conflicts of Interest

None declared.

Multimedia Appendix 1

HIWQ items.

[PDF File (Adobe PDF File), 21KB - jmir_v15i7e132_app1.pdf]

References

1. Berg JW, Appelbaum PS, Lidz CW, Parker LS. Informed consent: legal theory and clinical practice. Oxford: Oxford University Press; 2001.
2. Sheridan SL, Harris RP, Woolf SH, Shared Decision-Making Workgroup of the U.S. Preventive Services Task Force. Shared decision making about screening and chemoprevention. a suggested approach from the U.S. Preventive Services Task Force. Am J Prev Med 2004 Jan;26(1):56-66. [Medline: 14700714]
3. Committee on Quality of Health Care in America, Institute of Medicine. Crossing the quality chasm: a new health system for the 21st century. Washington, D.C: National Academy Press; 2001.
4. Lee YY, Lin JL. Do patient autonomy preferences matter? Linking patient-centered care to patient-physician relationships and health outcomes. Soc Sci Med 2010 Nov;71(10):1811-1818. [doi: 10.1016/j.socscimed.2010.08.008] [Medline: 20933316]
5. Bertakis KD, Azari R. Patient-centered care is associated with decreased health care utilization. J Am Board Fam Med 2011;24(3):229-239 [FREE Full text] [doi: 10.3122/jabfm.2011.03.100170] [Medline: 21551394]
6. Veroff D, Marr A, Wennberg DE. Enhanced support for shared decision making reduced costs of care for patients with preference-sensitive conditions. Health Aff (Millwood) 2013 Feb;32(2):285-293. [doi: 10.1377/hlthaff.2011.0941] [Medline: 23381521]
7. Parsi K, Chambers CJ, Armstrong AW. Cost-effectiveness analysis of a patient-centered care model for management of psoriasis. J Am Acad Dermatol 2012 Apr;66(4):563-570. [doi: 10.1016/j.jaad.2011.02.022] [Medline: 21835497]
8. Ballard-Reisch DS. A Model of Participative Decision Making for Physician-Patient Interaction. Health Communication 1990 Apr;2(2):91-104. [doi: 10.1207/s15327027hc0202_3]
9. McNutt RA. Shared medical decision making: problems, process, progress. JAMA 2004 Nov 24;292(20):2516-2518. [doi: 10.1001/jama.292.20.2516] [Medline: 15562133]
10. Brody DS. The patient's role in clinical decision-making. Ann Intern Med 1980 Nov;93(5):718-722. [Medline: 7212484]
11. Charles C, Gafni A, Whelan T. Shared decision-making in the medical encounter: what does it mean? (or it takes at least two to tango). Soc Sci Med 1997 Mar;44(5):681-692. [Medline: 9032835]

XSL•FO

RenderX

12. Charles C, Gafni A, Whelan T. Decision-making in the physician-patient encounter: revisiting the shared treatment decision-making model. Soc Sci Med 1999 Sep;49(5):651-661. [Medline: 10452420]

13. Braddock CH, Edwards KA, Hasenberg NM, Laidley TL, Levinson W. Informed decision making in outpatient practice: time to get back to basics. JAMA 1999;282(24):2313-2320. [Medline: 10612318]

14. Politi MC, Wolin KY, Légaré F. Implementing clinical practice guidelines about health promotion and disease prevention through shared decision making. J Gen Intern Med 2013 Jun;28(6):838-844. [doi: 10.1007/s11606-012-2321-0] [Medline: 23307397]

15. Entwistle VA, Watt IS. Patient involvement in treatment decision-making: the case for a broader conceptual framework. Patient Educ Couns 2006 Nov;63(3):268-278. [doi: 10.1016/j.pec.2006.05.002] [Medline: 16875797]

16. Ende J, Kazis L, Ash A, Moskowitz MA. Measuring patients' desire for autonomy: decision making and information-seeking preferences among medical patients. J Gen Intern Med 1989;4(1):23-30. [Medline: 2644407]

17. Xie B, Wang M, Feldman R. Preferences for health information and decision-making: development of the Health Information Wants questionnaire. In: Proceedings of the 2011 iConference. New York, NY: ACM; 2011 Presented at: iConference; Feb. 8-11, 2011; Seattle, Washington p. 237-280. [doi: 10.1145/1940761.1940799]

18. Krantz DS, Baum A, Wideman M. Assessment of Preferences for self-treatment and information in health care. J Pers Soc Psychol 1980 Nov;39(5):977-990. [Medline: 7441487]

19. Cassileth BR, Zupkis RV, Sutton-Smith K, March V. Information and participation preferences among cancer patients. Ann Intern Med 1980 Jun;92(6):832-836. [Medline: 7387025]

20. Galloway SC, Graydon JE. Uncertainty, symptom distress, and information needs after surgery for cancer of the colon. Cancer Nurs 1996 Apr;19(2):112-117. [Medline: 8635164]

21. Galloway S, Graydon J, Harrison D, Evans-Boyden B, Palmer-Wickham S, Burlein-Hall S, et al. Informational needs of women with a recent diagnosis of breast cancer: development and initial testing of a tool. J Adv Nurs 1997 Jun;25(6):1175-1183. [Medline: 9181414]

22. Chalmers KI, Luker KA, Leinster SJ, Ellis I, Booth K. Information and support needs of women with primary relatives with breast cancer: development of the Information and Support Needs Questionnaire. J Adv Nurs 2001 Aug;35(4):497-507. [Medline: 11529948]

23. Beisecker AE, Beisecker TD. Patient information-seeking behaviors when communicating with doctors. Med Care 1990 Jan;28(1):19-28. [Medline: 2296214]

24. Luker KA, Beaver K, Leinster SJ, Owens RG, Degner LF, Sloan JA. The information needs of women newly diagnosed with breast cancer. J Adv Nurs 1995 Jul;22(1):134-141. [Medline: 7560521]

25. Degner LF, Sloan JA. Decision making during serious illness: what role do patients really want to play? J Clin Epidemiol 1992 Sep;45(9):941-950. [Medline: 1432023]

26. Baker L, Wagner TH, Singer S, Bundorf MK. Use of the Internet and e-mail for health care information: results from a national survey. JAMA 2003 May 14;289(18):2400-2406. [doi: 10.1001/jama.289.18.2400] [Medline: 12746364]

27. Fox S. 2006 Online Health Search.: Pew Internet Research URL: http://www.pewinternet.org/~/media/Files/Reports/2006/PIP_Online_Health_2006.pdf.pdf [accessed 2013-02-17] [WebCite Cache ID 6EVzd44g1]

28. Fox S, Raine L. Vital decisions: how Internet users decide what information to trust when they or their loved ones are sick. Plus a guide from the Medical Library Association about smart health-search strategies and good Web sites. 2002. URL: http://www.pewinternet.org/~/media//Files/Reports/2002/PIP_Vital_Decisions_May2002.pdf.pdf [accessed 2013-02-17] [WebCite Cache ID 6EW1JiG7H]

29. Fox S. Health Information Online.: Pew Internet Research; 2005. URL: http://www.pewinternet.org/pdfs/PIP_Healthtopics_May05.pdf [accessed 2013-02-17] [WebCite Cache ID 6EW2x0kjg]

30. Fox S, Raine L. The online health care revolution: how the Web helps Americans take better care of themselves. 2000. URL: http://www.pewinternet.org/~/media//Files/Reports/2000/PIP_Health_Report.pdf.pdf [accessed 2013-02-18] [WebCite Cache ID 6EW33MwcI]

31. Dickerson S, Reinhart AM, Feeley TH, Bidani R, Rich E, Garg VK, et al. Patient Internet use for health information at three urban primary care clinics. J Am Med Inform Assoc 2004;11(6):499-504 [FREE Full text] [doi: 10.1197/jamia.M1460] [Medline: 15298993]

32. Rozmovits L, Ziebland S. What do patients with prostate or breast cancer want from an Internet site? A qualitative study of information needs. Patient Educ Couns 2004 Apr;53(1):57-64. [doi: 10.1016/S0738-3991(03)00116-2] [Medline: 15062905]

33. Kaiser Family Foundation. e-Health and the elderly: how seniors use the Internet for health information. 2005. URL: http://www.kff.org/entmedia/upload/e-Health-and-the-Elderly-How-Seniors-Use-the-Internet-for-Health-Information-Key-Findings-From-a-National-Survey-of-Older-Americans-Survey-Report.pdf [accessed 2013-02-18] [WebCite Cache ID 6EW3J0BOp]

34. Fox S. E-patients with a disability or chronic disease. Washington, DC: PEW Internet & Life Project; 2007. URL: http://www.pewinternet.org/~/media//Files/Reports/2007/EPatients_Chronic_Conditions_2007.pdf.pdf [accessed 2013-02-18] [WebCite Cache ID 6EW3NyUk5]

35. Greaves F, Pape UJ, King D, Darzi A, Majeed A, Wachter RM, et al. Associations between Internet-based patient ratings and conventional surveys of patient experience in the English NHS: an observational study. BMJ Qual Saf 2012 Jul;21(7):600-605. [doi: 10.1136/bmjqs-2012-000906] [Medline: 22523318]

36. Hu X, Bell RA, Kravitz RL, Orrange S. The prepared patient: information seeking of online support group members before their medical appointments. J Health Commun 2012;17(8):960-978. [doi: 10.1080/10810730.2011.650828] [Medline: 22574697]

37. Groves ND, Humphreys HW, Williams AJ, Jones A. Effect of informational internet web pages on patients' decision-making: randomised controlled trial regarding choice of spinal or general anaesthesia for orthopaedic surgery. Anaesthesia 2010 Mar;65(3):277-282. [Medline: 20336817]

38. Kassan EC, Williams RM, Kelly SP, Barry SA, Penek S, Fishman MB, et al. Men's use of an Internet-based decision aid for prostate cancer screening. J Health Commun 2012;17(6):677-697. [doi: 10.1080/10810730.2011.579688] [Medline: 21919646]

39. Ziebland S, Chapple A, Dumelow C, Evans J, Prinjha S, Rozmovits L. How the internet affects patients' experience of cancer: a qualitative study. BMJ 2004 Mar 6;328(7439):564 [FREE Full text] [doi: 10.1136/bmj.328.7439.564] [Medline: 15001506]

40. Xie B. Older adults' health information wants in the internet age: implications for patient-provider relationships. J Health Commun 2009 Sep;14(6):510-524. [doi: 10.1080/10810730903089614] [Medline: 19731124]

41. Xie B, Wang M, Feldman R, Zhou L. Health information and decision-making preferences in the Internet age: a pilot study using the Health Information Wants (HIW) Questionnaire. In: Proceedings of the 1st ACM International Health Informatics Symposium. New York, NY: ACM; 2010 Presented at: 1st ACM International Health Informatics Symposium; Nov. 11-12, 2010; Arlington, VA p. 610-619. [doi: 10.1145/1882992.1883090]

42. Xie B, Wang M, Feldman R, Zhou L. Exploring older and younger adults' preferences for health information and participation in decision making using the Health Information Wants Questionnaire (HIWQ). Health Expect 2012 Jul 11:---. [doi: 10.1111/j.1369-7625.2012.00804.x] [Medline: 22783800]

43. Zickuhr K, Madden A. Older adults and Internet use.: Pew Internet Research; 2012. URL: http://www.pewinternet.org/~/media//Files/Reports/2012/PIP_Older_adults_and_internet_use.pdf [accessed 2013-02-18] [WebCite Cache ID 6EW44YZMW]

44. Levinson W, Kao A, Kuby A, Thisted RA. Not all patients want to participate in decision making. A national study of public preferences. J Gen Intern Med 2005 Jun;20(6):531-535 [FREE Full text] [doi: 10.1111/j.1525-1497.2005.04101.x] [Medline: 15987329]

45. Pinquart M, Duberstein PR. Information needs and decision-making processes in older cancer patients. Crit Rev Oncol Hematol 2004 Jul;51(1):69-80. [doi: 10.1016/j.critrevonc.2004.04.002] [Medline: 15207255]

46. Benbassat J, Pilpel D, Tidhar M. Patients' preferences for participation in clinical decision making: a review of published surveys. Behav Med 1998;24(2):81-88. [doi: 10.1080/08964289809596384] [Medline: 9695899]

47. Hill SA, Laugharne R. Decision making and information seeking preferences among psychiatric patients. J Ment Health 2006 Jan;15(1):75-84. [doi: 10.1080/09638230500512250]

48. Flynn KE, Smith MA. Personality and health care decision-making style. J Gerontol B Psychol Sci Soc Sci 2007 Sep;62(5):P261-P267 [FREE Full text] [Medline: 17906167]

49. Auerbach SM, Pegg PO. Appraisal of Desire for Control over Healthcare: Structure, Stability, and Relation to Health Locus of Control and to the 'Big Five' Personality Traits. J Health Psychol 2002 Jul;7(4):393-408. [doi: 10.1177/1359105302007004328] [Medline: 22112750]

50. Braman AC, Gomez RG. Patient personality predicts preference for relationships with doctors. Personality and Individual Differences 2004 Sep;37(4):815-826. [doi: 10.1016/j.paid.2003.10.009]

51. Bagley-Burnett C. Measuring information-seeking behaviorsdecision-making preferences. In: Frank-Stromborg M, Olsen SJ, editors. Instruments for Clinical Health-Care Research 3rd ed. Sudbury, MA: Jones and Bartlett Publishers; 2004:455-471.

52. Say R, Murtagh M, Thomson R. Patients' preference for involvement in medical decision making: a narrative review. Patient Educ Couns 2006 Feb;60(2):102-114. [doi: 10.1016/j.pec.2005.02.003] [Medline: 16442453]

53. Davis RE, Jacklin R, Sevdalis N, Vincent CA. Patient involvement in patient safety: what factors influence patient participation and engagement? Health Expect 2007 Sep;10(3):259-267. [doi: 10.1111/j.1369-7625.2007.00450.x] [Medline: 17678514]

54. Jung HP, Baerveldt C, Olesen F, Grol R, Wensing M. Patient characteristics as predictors of primary health care preferences: a systematic literature analysis. Health Expect 2003 Jun;6(2):160-181. [Medline: 12752744]

55. Chewning B, Bylund CL, Shah B, Arora NK, Gueguen JA, Makoul G. Patient preferences for shared decisions: a systematic review. Patient Educ Couns 2012 Jan;86(1):9-18. [doi: 10.1016/j.pec.2011.02.004] [Medline: 21474265]

56. Juster F, Suzman R. The health and retirement study: an overview. J Hum Resour 1995;30(Suppl):7-56.

57. Raudenbush SW, Bryk AS. Hierarchical linear models: Applications and data analysis methods, 2nd ed. Thousand Oaks, CA: Sage; 2002.

58. Blumenthal D. Doctors in a wired world: can professionalism survive connectivity? Milbank Q 2002;80(3):525-46, iv [FREE Full text] [Medline: 12233248]

59.   Gerber BS, Eiser AR. The patient physician relationship in the Internet age: future prospects and the research agenda. J Med Internet Res 2001;3(2):E15 [FREE Full text] [doi: 10.2196/jmir.3.2.e15] [Medline: 11720957]

60.   Wald HS, Dube CE, Anthony DC. Untangling the Web--the impact of Internet use on health care and the physician-patient relationship. Patient Educ Couns 2007 Nov;68(3):218-224. [doi: 10.1016/j.pec.2007.05.016] [Medline: 17920226]

61.   Rozenblum R, Bates DW. Patient-centred healthcare, social media and the internet: the perfect storm? BMJ Qual Saf 2013 Mar;22(3):183-186. [doi: 10.1136/bmjqs-2012-001744] [Medline: 23378660]

62.   Starr P. The social transformation of American medicine. New York: Basic Books, Inc., Publishers; 1982.

63.   Fox S, Duggan M. Health Online 2013.: Pew Internet Research URL: http://pewinternet.org/~/media/ /Files/Reports/PIP_HealthOnline.pdf [accessed 2013-02-17] [WebCite Cache ID 6EVyj3wus]

64.   Lima-Pereira P, Bermúdez-Tamayo C, Jasienska G. Use of the Internet as a source of health information amongst participants of antenatal classes. J Clin Nurs 2012 Feb;21(3-4):322-330. [doi: 10.1111/j.1365-2702.2011.03910.x] [Medline: 22093043]

65.   Campbell RJ. Consumer Health, Patient Education, and The Internet. Internet Journal of Health 2001;2(2):---. [doi: 10.5580/289e#sthash.SM9hihV8.dpuf]

66.   Hardt JH, Hollis-Sawyer L. Older Adults Seeking Healthcare Information on the Internet. Educational Gerontology 2007 Jun 13;33(7):561-572. [doi: 10.1080/03601270701364628]

67.   Greene JA, Choudhry NK, Kilabuk E, Shrank WH. Online social networking by patients with diabetes: a qualitative evaluation of communication with Facebook. J Gen Intern Med 2011 Mar;26(3):287-292 [FREE Full text] [doi: 10.1007/s11606-010-1526-3] [Medline: 20945113]

68.   Greaves F, Millett C. Consistently increasing numbers of online ratings of healthcare in England. J Med Internet Res 2012;14(3):e94 [FREE Full text] [doi: 10.2196/jmir.2157] [Medline: 22742977]

69.   Gao GG, McCullough JS, Agarwal R, Jha AK. A changing landscape of physician quality reporting: analysis of patients' online ratings of their physicians over a 5-year period. J Med Internet Res 2012 Feb;14(1):e38 [FREE Full text] [doi: 10.2196/jmir.2003] [Medline: 22366336]

70.   Bardach NS, Asteria-Peñaloza R, Boscardin WJ, Dudley RA. The relationship between commercial website ratings and traditional hospital performance measures in the USA. BMJ Qual Saf 2013 Mar;22(3):194-202. [doi: 10.1136/bmjqs-2012-001360] [Medline: 23178860]

71.   Greaves F, Ramirez-Cano D, Millett C, Darzi A, Donaldson L. Harnessing the cloud of patient experience: using social media to detect poor quality healthcare. BMJ Qual Saf 2013 Mar;22(3):251-255. [doi: 10.1136/bmjqs-2012-001527] [Medline: 23349387]

72.   Kaplan RM, Frosch DL. Decision making in medicine and health care. Annu Rev Clin Psychol 2005;1:525-556. [doi: 10.1146/annurev.clinpsy.1.102803.144118] [Medline: 17716098]

73.   Gaston CM, Mitchell G. Information giving and decision-making in patients with advanced cancer: a systematic review. Soc Sci Med 2005 Nov;61(10):2252-2264. [doi: 10.1016/j.socscimed.2005.04.015] [Medline: 15922501]

74.   Deber RB, Kraetschmer N, Irvine J. What role do patients wish to play in treatment decision making? Arch Intern Med 1996 Jul 8;156(13):1414-1420. [Medline: 8678709]

75.   Strull WM, Lo B, Charles G. Do patients want to participate in medical decision making? JAMA 1984 Dec 7;252(21):2990-2994. [Medline: 6502860]

76.   Stevenson LW, Hellkamp AS, Leier CV, Sopko G, Koelling T, Warnica JW, et al. Changing preferences for survival after hospitalization with advanced heart failure. J Am Coll Cardiol 2008 Nov 18;52(21):1702-1708 [FREE Full text] [doi: 10.1016/j.jacc.2008.08.028] [Medline: 19007689]

77.   Bowling A, Reeves BC, Rowe G. Do treatment preferences for patients with angina change? An 18-month follow-up study. Health Expect 2012 Dec;15(4):351-359 [FREE Full text] [doi: 10.1111/j.1369-7625.2011.00680.x] [Medline: 21726362]

78.   Pardon K, Deschepper R, Vander Stichele R, Bernheim JL, Mortier F, Bossuyt N, EOLIC-consortium. Changing preferences for information and participation in the last phase of life: a longitudinal study among newly diagnosed advanced lung cancer patients. Support Care Cancer 2012 Oct;20(10):2473-2482. [doi: 10.1007/s00520-011-1369-4] [Medline: 22246616]

79.   Loukanova S, Molnar R, Bridges JF. Promoting patient empowerment in the healthcare system: highlighting the need for patient-centered drug policy. Expert Rev Pharmacoecon Outcomes Res 2007 Jun;7(3):281-289. [doi: 10.1586/14737167.7.3.281] [Medline: 20528314]

Abbreviations

CAM: complementary and alternative medicine
HIW: Health Information Wants
HIWQ: Health Information Wants Questionnaire

XSL•FO
RenderX

Review

Systematic Reviews and Meta-Analyses of Home Telemonitoring Interventions for Patients With Chronic Diseases: A Critical Assessment of Their Methodological Quality

Spyros Kitsiou[1], PhD; Guy Paré[1*], PhD; Mirou Jaana[2,3*], PhD

[1]Canada Research Chair in Information Technology in Health Care, HEC Montreal, Montreal, QC, Canada

[2]Telfer School of Management, University of Ottawa, Ottawa, ON, Canada

[3]School of Business, Lebanese American University, Beirut, Lebanon

[*]these authors contributed equally

Corresponding Author:
Spyros Kitsiou, PhD
HEC Montreal
3000, chemin de la Côte-Sainte-Catherine
Montreal, QC, H3T 2A7
Canada
Phone: 1 514 340 6000 ext 2653
Fax: 1 514 340 6132
Email: spyros.kitsiou@hec.ca

Related Article:

This is a corrected version. See correction statement: http://www.jmir.org/2013/11/e253/

Abstract

Background: Systematic reviews and meta-analyses of home telemonitoring interventions for patients with chronic diseases have increased over the past decade and become increasingly important to a wide range of clinicians, policy makers, and other health care stakeholders. While a few criticisms about their methodological rigor and synthesis approaches have recently appeared, no formal appraisal of their quality has been conducted yet.

Objective: The primary aim of this critical review was to evaluate the methodology, quality, and reporting characteristics of prior reviews that have investigated the effects of home telemonitoring interventions in the context of chronic diseases.

Methods: Ovid MEDLINE, the Database of Abstract of Reviews of Effects (DARE), and Health Technology Assessment Database (HTA) of the Cochrane Library were electronically searched to find relevant systematic reviews, published between January 1966 and December 2012. Potential reviews were screened and assessed for inclusion independently by three reviewers. Data pertaining to the methods used were extracted from each included review and examined for accuracy by two reviewers. A validated quality assessment instrument, R-AMSTAR, was used as a framework to guide the assessment process.

Results: Twenty-four reviews, nine of which were meta-analyses, were identified from more than 200 citations. The bibliographic search revealed that the number of published reviews has increased substantially over the years in this area and although most reviews focus on studying the effects of home telemonitoring on patients with congestive heart failure, researcher interest has extended to other chronic diseases as well, such as diabetes, hypertension, chronic obstructive pulmonary disease, and asthma. Nevertheless, an important number of these reviews appear to lack optimal scientific rigor due to intrinsic methodological issues. Also, the overall quality of reviews does not appear to have improved over time. While several criteria were met satisfactorily by either all or nearly all reviews, such as the establishment of an a priori design with inclusion and exclusion criteria, use of electronic searches on multiple databases, and reporting of studies characteristics, there were other important areas that needed improvement. Duplicate data extraction, manual searches of highly relevant journals, inclusion of gray and non-English literature, assessment of the methodological quality of included studies and quality of evidence were key methodological procedures that were performed infrequently. Furthermore, certain methodological limitations identified in the synthesis of study results have affected the results and conclusions of some reviews.

Conclusions: Despite the availability of methodological guidelines that can be utilized to guide the proper conduct of systematic reviews and meta-analyses and eliminate potential risks of bias, this knowledge has not yet been fully integrated in the area of home telemonitoring. Further efforts should be made to improve the design, conduct, reporting, and publication of systematic reviews and meta-analyses in this area.

Introduction

The prevalence of chronic diseases such as diabetes, cardiovascular, and respiratory conditions continues to pose a significant and longstanding challenge for virtually all health care systems, requiring fundamental changes in the management and delivery of patient care [1-3]. Home telemonitoring (HT) represents a promising approach for enabling patients with chronic conditions to be followed up by clinicians more frequently, over longer periods of time, away from hospital settings [4-6]. HT is a particular form of telehealth that encompasses the use of remote access information and communication technologies (eg, telemetry devices, intelligent sensors, hand-held or wearable technologies) for the timely transmission of symptoms, physiological, and disease-related data from the patients' home to a telemonitoring center supporting clinical decisions [4,5,7]. The underlying goal of HT is to provide doctors and nurses with accurate and timely information necessary to remotely detect any abnormal health parameters and complications associated with the disease, earlier than during a scheduled follow-up or an emergency visit. This allows timely interventions before exacerbations and complications occur, necessitating admission to the hospital and use of more resources.

Over the years, in the context of national eHealth strategies in Europe, Canada, Australia, the United States, and other parts of the world, there have been numerous efforts and research initiatives to examine the effectiveness of HT for patients with chronic diseases as a potential cost-saving approach (eg, [8,9]). The Veterans Health Administration's extensive home telehealth service in the United States [10] and the Whole System Demonstrator (WSD) program in the United Kingdom [11] are a few examples. Nonetheless, the benefits from wider diffusion and use of HT applications have not been fully achieved yet [12]. The confidence and acceptance of health authorities to support and reimburse HT services for the management of chronic diseases depend to a large extent on the availability of reliable and robust scientific evidence from the field [13].

Systematic reviews (SRs) and meta-analyses (MAs) are powerful research tools that have been established in the health sciences, and more recently in the medical informatics field, as the cornerstone of evidence-based practice [14,15]. They adhere closely to a set of rigid scientific guidelines and use rigorous and reproducible methods to identify, select, appraise, and synthesize the results of clinical studies, in order to minimize the potential for bias in addressing a specific research question [16]. SRs and MAs have become increasingly important in the health care domain and their value to policy makers, clinicians, and researchers is well recognized [17]. When properly conducted, they provide relevant information for policy makers and serve as the foundation for the development of evidence-based practice and clinical guidelines.

However, the quality and internal validity of SRs and MAs depend on many aspects pertaining to the conduct of the review and the quality of empirical studies selected for inclusion. Flaws and deficiencies in the methods concerning the bibliographic search, selection, appraisal, and synthesis of evidence can lead to invalid conclusions with significant implications for patient care and decision makers. Hence, researchers have proposed and adopted evaluation tools that allow a close examination of the methodological rigor of reviews in several clinical areas (eg, [17-21]).

Reviews focusing on HT interventions for patients with chronic diseases have increased over the past decade. While a few criticisms about their methodological rigor and approaches have recently appeared (eg, [6,11,22,23]), no formal appraisal of their scientific quality has been conducted yet. This paper attempts to fill this gap by evaluating the methodology, quality, and reporting characteristics of SRs and MAs of HT interventions in the context of chronic diseases, in order to identify risks of bias that may have affected their internal validity. In studying and presenting methodological deficiencies identified in prior reviews, we do not intend to exemplify author incompetence. In fact, many of the authors of the included reviews are rightly acknowledged as leading experts and most of the included papers have provided the base for building evidence in a relatively recent discipline. However, we truly believe that scientific progress in this particular area of HT will not occur through the accumulation of uncontested findings, but through a continuous process of constructive criticism, vigorous debate, and creation of awareness [24]. To this end, our objective is to constructively inform other scholars and strengthen knowledge development by giving focus and direction to future reviews of HT for further improvement.

Methods

Inclusion and Exclusion Criteria

Overview

All inclusion and exclusion criteria were defined a priori. Citations identified in the search were assessed for eligibility against the study selection criteria explained below: types of studies, patients, interventions, and outcomes.

Types of Studies

Only prior SRs and MAs considering the effects of HT and published in peer-reviewed journals or the Cochrane Library were eligible for inclusion. To determine during the screening process whether a published article corresponded to these review types, we relied on key characteristics outlined by the Cochrane Collaboration [25]. In particular, we considered a review to be systematic if it included: (1) a set of clearly formulated research objectives or research questions with predetermined eligibility criteria for the selection of relevant empirical studies, (2) an explicit, reproducible methodology, (3) a systematic search strategy that attempted to identify all studies that would meet the eligibility criteria, and (4) a systematic presentation, analysis, and synthesis of the characteristics and findings of the included studies. Depending on the methods used to summarize and synthesize the available evidence from primary studies, systematic reviews can be classified as qualitative/narrative or quantitative (ie, meta-analyses). In our sample we included both MAs and narrative SRs. Reviews that were self-described as systematic, whether in the title, abstract, or methods of the paper, were also included. These criteria were utilized regardless of the quality or comprehensiveness of the review. We excluded conference proceedings, review summaries, editorials, and unpublished works.

Types of Patients and Interventions

In order to meet the inclusion criteria, the reviews had to investigate the effectiveness of HT interventions for patients with one of the following chronic conditions: congestive heart failure, hypertension, diabetes, or respiratory conditions. They also had to include primary (empirical) studies that involved the use of information and communication technologies by patients for the timely transmission and remote monitoring of vital signs (eg, arterial blood pressure, cardiac rate), biometric, and disease-related data (eg, blood glucose levels, symptoms, use of medication) from the patients' residence to a clinician (eg, nurse, doctor, or allied health professional) at a health care service center. SRs that investigated and combined collectively (ie, without making a distinction) the effects of HT with other stand-alone multidisciplinary interventions of remote patient monitoring (eg, structured telephone support, telediagnosis, or teleconsultation) were excluded.

Outcomes

Prior reviews were included only if primary or secondary outcomes from the primary studies pertaining to the clinical, structural (eg, utilization of services), behavioral (eg, impacts on patients' behavior), or economic effects of HT were synthesized and presented. Reviews that focused on other aspects such as the technical feasibility of HT modalities were excluded.

Search Strategy

We performed a literature search on Ovid MEDLINE, the Database of Abstract of Reviews of Effects (DARE), and Health Technology Assessment Database (HTA) of the Cochrane Library (from 1966 to December 2012) in order to identify all relevant reviews. On the Cochrane Library, we conducted the search using four keywords (telemonitoring, telecare, telehealth,

telehomecare). On Ovid MEDLINE, we used the same keywords in conjunction with each of the following terms: systematic review, meta-analysis, and review. Language restrictions were not applied to any of the searches.

Selection of Relevant Reviews

As shown in Figure 1, our initial search resulted in 240 references after eliminating duplicates. The title and abstract of these references were examined independently by the 3 authors to identify articles that appeared potentially relevant to this study area. Any differences were resolved by discussion until consensus was achieved. Based on the inclusion criteria, 185 references were deemed not relevant and were excluded. The remaining 55 were identified as potentially relevant, and full copies of these references were retrieved for further assessment. The reference lists of these articles were manually examined to identify potentially relevant reviews that were not originally captured in the initial search. This process yielded 16 additional references. Several reviews were excluded as they concerned other forms of telehealth interventions (n=24), they included primary studies with multipathology patients (n=8) or reviewed topics other than the effectiveness of HT (n=2). Other studies were excluded because they were not SRs or MAs (n=10), and 2 reviews were excluded as they were published in a language other than English. Multimedia Appendix 1 provides the full list of references that were excluded. The final number of SRs included in this critical review was 24 [26-49]. Note that one review was published initially as a Cochrane Collaboration review [31], and later an abridged version of it appeared in a journal [50]. In our assessment, we used the former publication as it is more detailed.

Extraction of Information

One reviewer (SK) extracted explicit details from each review in a nonblinded manner by using an electronic extraction form that was developed for the purposes of this study. All extracted data were examined for accuracy by 2 of the reviewers (GP and MJ), and any disagreements were reconciled through consensus. The information sought included general details pertaining to the characteristics of the reviews (eg, number of authors, origin of the corresponding author, year of publication, journal characteristics, sources of funding) and more specific details about the use and interpretation of methods for synthesizing the available evidence (eg, meta-analytic and qualitative techniques).

Assessment of Methodological Quality

The methodological quality of the 24 reviews was appraised independently in a nonblinded format by 2 reviewers (SK and GP) using the Revised Assessment of Multiple Systematic Reviews (R-AMSTAR) instrument [51]. Any disagreements were reconciled through consensus. R-AMSTAR was chosen on the basis that it is a validated instrument that offers the ability to conduct an in-depth appraisal of SRs and MAs by assessing the presence of (1) an a priori design, (2) duplicate study selection and data extraction, (3) a comprehensive literature search, (4) the inclusion of gray literature, (5) a list of included/excluded studies, (6) a profile of the included studies, (7) a documented assessment of the scientific quality of included

studies, (8) the appropriate use of the scientific quality in forming conclusions, (9) the appropriate use of methods to combine findings of studies, (10) the assessment of the likelihood of publication bias, and (11) the proper documentation of conflict of interest. Each of these domains will be described in greater detail later.

Figure 1. Flow diagram describing the selection process of SRs and MAs.
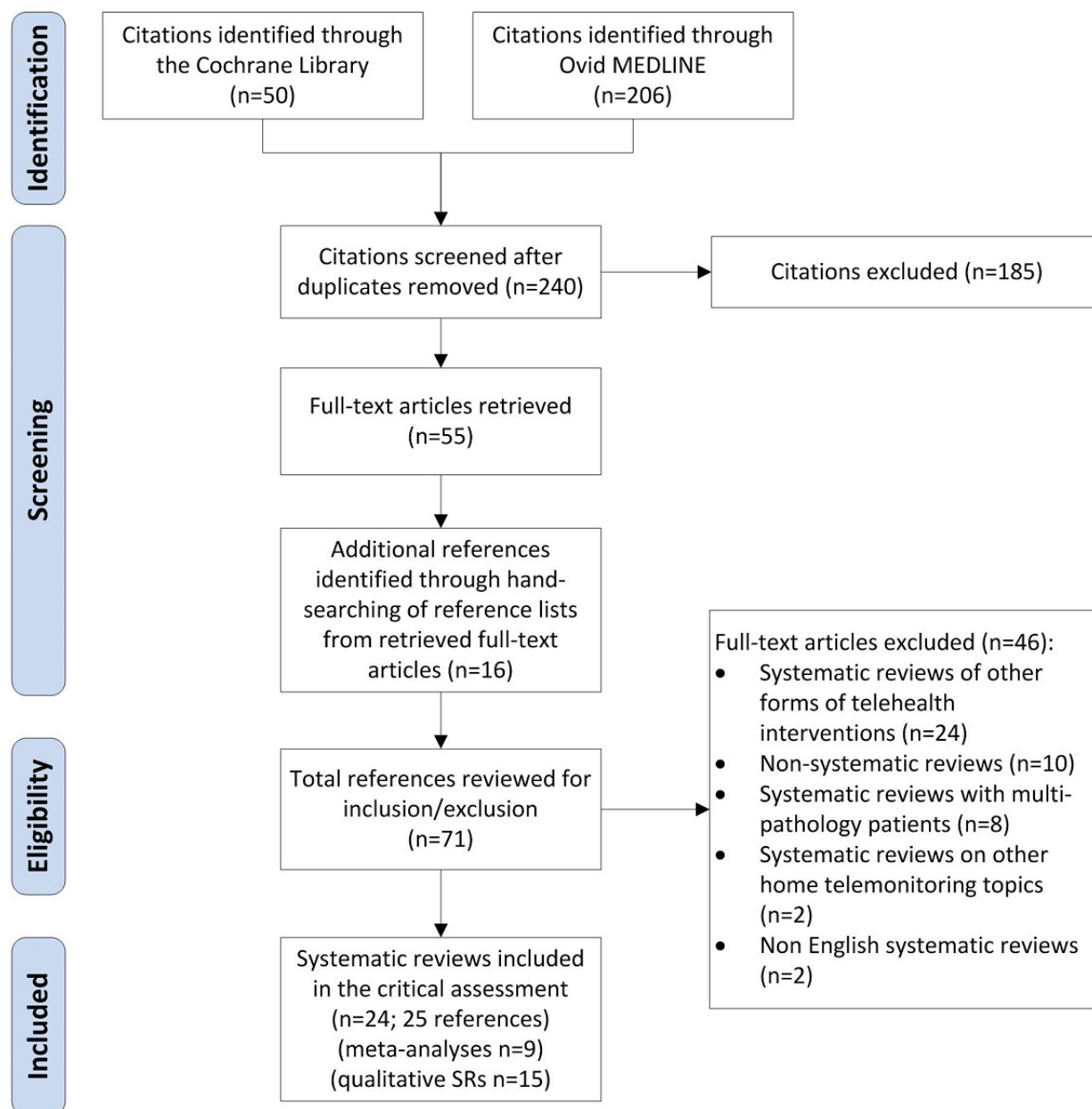
Results

Profile of the Reviews

Figure 2 displays the trend over time in the publication of SRs and MAs of HT interventions. Our findings reveal that the first review was published in 2003 [32]. Clearly, very few reviews were published prior to 2007. But since then, the number of HT reviews has increased substantially.

As shown in Table 1, the largest body of reviews (n=10) focused on the effects of HT on patients with congestive heart failure [26-35]. Four reviews (17%) considered patients with hypertension [36-39]; 4 reviews (17%) examined HT for patients with respiratory conditions such as chronic obstructive pulmonary disease (n=2), cystic fibrosis (n=1), and asthma (n=1) [40-43]; and 4 other reviews (17%) focused on patients with diabetes [44-47]. Last, our sample comprises 2 comprehensive SRs (8%), which investigated the effects of HT across various chronic diseases (ie, heart failure, hypertension, diabetes, and respiratory conditions) [48,49]. These reviews were included since HT effects were reported separately for each condition.

All but 3 reviews were published in peer-reviewed journals. The 3 most common sources were the *Journal of Telemedicine and Telecare* (n=3), *Telemedicine and e-Health* (n=3), and the

Journal of Evaluation in Clinical Practice (n=2). Five reviews [27,31,42,46,48] reported being updates of previous reviews. In most articles, the corresponding authors were from North America with 10 being from Canada and 4 from the United States. Six reviews originated in Europe (4 in the United Kingdom, 1 in Greece, and 1 in Italy), 3 in Australia, and 1 in Taiwan. Six reviews comprised a multinational group of researchers.

Most reviews were conducted by 2 or more authors and only 2 [35,42] were single authored. The majority of reviews (63%) were funded by government organizations or health care agencies. Five of these received additional funding either from the industry or from academic institutions. Less than half of the reviews combined the results from the primary studies into an

MA, and most reviews (63%) used qualitative approaches to synthesize the available evidence. MAs were found to be cited more frequently (mean 103.6, SD 108.2, 95% CI 13.1-194.1) than SRs (mean 61.1, SD 77.2, 95% CI 18.37-103.90), but this difference was not statistically significant (P=.287).

Methodological Quality of Reviews

The results of the methodological quality of the included reviews are presented in Table 2. We outline all 41 quality criteria covered by the R-AMSTAR instrument and present the percentage of review articles that met each of them. Multimedia Appendix 2 provides a detailed analysis of each review. We list in lower-case letters all the criteria that were covered satisfactorily [51]. In the following sections, we present an analysis of the key findings within each R-AMSTAR domain.

Figure 2. Number of HT systematic reviews and meta-analyses published per year.
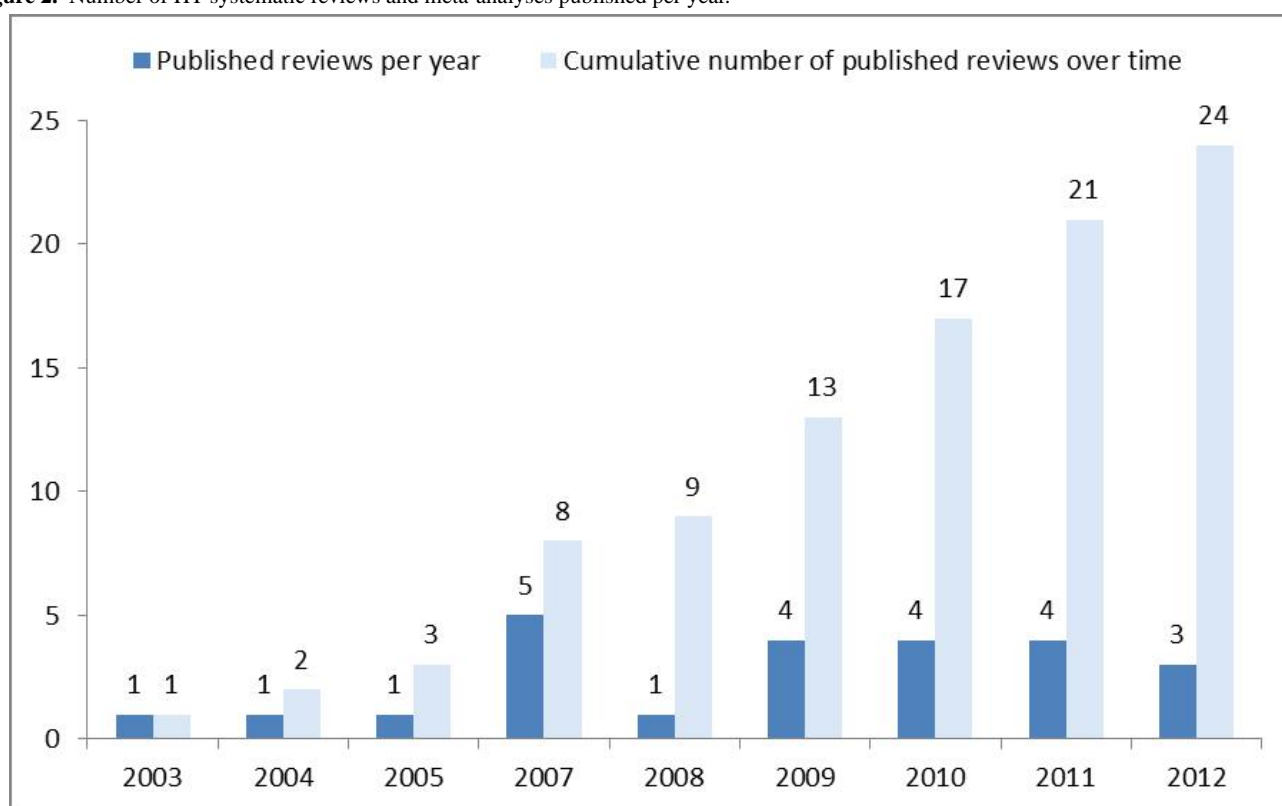
Table 1. Profile of the reviews.

Chronic disease	Reference	Year	Type of Review	Number of cites[a]	Period covered	Total # of included studies (number of RCTs[d])
Heart failure						
	Chaudhry et al [26]	2007	SR[b]	94	1966-2006	9 (9)
	Clark et al [27]	2007	MA[c]	323	2002-2006	5 (5)
	Clarke et al [28]	2011	MA	23	1969-2009	13 (13)
	Dang et al [29]	2009	SR	30	1966-2009	9 (9)
	Giamouzis et al [30]	2012	SR	4	2001-2011	12 (12)
	Inglis et al [31]	2010	MA	173	2006-2008	14 (14)
	Louis et al [32]	2003	SR	199	1966-2002	24 (6)
	Maric et al [33]	2009	SR	53	up to 2007	41 (12)
	Polisena et al [34]	2010	MA	50	1998-2008	21 (11)
	Seto [35]	2008	SR	48	up to 2007	8 (4)
Hypertension						
	AbuDagga et al [36]	2010	SR	18	1995-2009	15 (10)
	Jaana et al [37]	2007	SR	13	1966-2006	14 (3)
	Omboni et al [38]	2011	MA	7	up to 2010	12 (12)
	Verberk et al [39]	2011	MA	6	not reported	9 (9)
Respiratory conditions						
	Bolton et al [40]	2011	SR	16	1990-2009	6 (2)
	Cox et al [41]	2012	SR	1	1998-2011	8 (1)
	Franek et al [42]	2012	SR	4	2000-2010	5 (3)
	Jaana et al [43]	2009	SR	49	1966-2007	14 (3)
Diabetes						
	Farmer et al [44]	2005	MA	127	1966-2004	26 (16)
	Jaana et al [45]	2007	SR	70	not reported	17 (11)
	MAS [46]	2009	MA	-	2007-2009	8 (8)
	Montori et al [47]	2004	MA	120	1982-2003	8 (8)
SRs covering various chronic diseases						
	Paré et al [48]	2010	SR	44	1966-2008	CHF: 17 (13); Hypertension: 13 (5); Asthma: 8 (6); Diabetes: 24 (21)
	Paré et al [49]	2007	SR	274	1990-2006	CHF: 16 (7); Hypertension: 14 (3); Respiratory Conditions: 18 (4); Diabetes: 17 (12)

[a]According to *Google Scholar* as of March 28, 2013.

[b]SR: Narrative/Qualitative systematic review.

[c]MA: Meta-analysis.

[d]randomized controlled trials.

Table 2. Percentage of reviews that satisfactorily met each R-AMSTAR criterion.

Criterion	Description	Yes, %
Q 1.a	The design of the study was established before the conduct of the review (ie, a priori design).	100
Q 1.b	There was a statement of inclusion criteria.	100
Q 1.c	There was a PICO research question/statement.	67
Q 2.a	There were at least 2 independent data extractors as stated or implied.	42
Q 2.b	There was a statement of recognition or awareness of consensus procedure for disagreements.	46
Q 2.c	Disagreements among extractors were resolved properly as stated or implied.	38
Q 3.a	At least 2 electronic sources were searched (eg, Medline and EMBASE).	96
Q 3.b	The report includes years and databases searched.	92
Q 3.c	Key words and/or MESH terms are stated.	92
Q 3.d	In addition to the electronic databases (PubMed, EMBASE, Medline), the search was supplemented by consulting current contents such as reviews, textbooks, specialized registers, or experts in the particular field of study or by reviewing the references in the studies found.	79
Q 3.e	Journals were "hand searched" or "manual searched" (ie, identifying highly relevant journals and conducting a manual, page-by-page search of their entire contents looking for potentially eligible studies).	13
Q 4.a	The authors stated that they searched for reports regardless of publication type.	8
Q 4.b	The authors state whether or not they excluded any reports (from the systematic review), based on their publication status, language, etc.	83
Q 4.c	"NonEnglish" papers were translated.	4
Q 4.d	There was no language restriction or recognition of nonEnglish articles.	21
Q 5.a	Table/list/or figure of included studies was provided; a reference list does not suffice.	92
Q 5.b	Table/list/or figure of excluded studies was provided either in the article or in a supplemental source (ie, online). (Excluded studies refers to those studies seriously considered on the basis of title and/or abstract, but rejected after reading the body of the text.)	25
Q 5.c	Author satisfactorily/sufficiently stated the reason for exclusion of the seriously considered studies.	63
Q 5.d	Reader is able to retrace the included and the excluded studies anywhere in the article bibliography, reference, or supplemental source.	25
Q 6.a	The characteristics of the included studies are provided in an aggregated form such as a table, data from the original studies were provided on the participants, interventions AND outcomes.	88
Q 6.b	The authors provided the ranges of relevant characteristics in the studies analyzed (eg, age, race, sex, relevant socioeconomic data, disease status, duration, severity, or other diseases are reported).	83
Q 6.c	The information provided appears to be complete and accurate (ie, there is a tolerable range of subjectivity here. Is the reader left wondering? If so, state the needed information and the reasoning).	88
Q 7.a	A priori methods of assessment were provided (eg, for effectiveness studies if the author(s) chose to include only randomized, double-blind, placebo controlled studies, or allocation concealment as inclusion criteria); for other types of studies alternative items will be relevant.	38
Q 7.b	The scientific quality of the included studies appears to be meaningful (ie, a scale such as High, Low or A, B, C is used).	33
Q 7.c	Discussion/recognition/awareness of level of evidence	21
Q 7.d	Quality of evidence was rated/ranked based on characterized instruments (Characterized instrument is a created instrument that ranks the level of evidence, eg, GRADE).	21
Q 8.a	The results of the methodological rigor and scientific quality were considered in the analysis and the conclusions of the SR.	25
Q 8.b	The results of the methodological rigor and scientific quality were explicitly stated in formulating recommendations.	25
Q 8.c	To have conclusions integrated/drives towards a clinical consensus statement.	n/a
Q 8.d	This clinical consensus statement drives toward revision or confirmation of clinical practice guidelines.	n/a
Q 9.a	The authors provided a statement of criteria that were used to decide that the studies analyzed were similar enough to be pooled.	0
Q 9.b	For the pooled results, a test was performed to ensure the studies were combinable, to assess their homogeneity (ie, Chi-square test for homogeneity, I^2).	38

Criterion	Description	Yes, %
Q 9.c	There was a recognition of heterogeneity or lack of thereof.	38
Q 9.d	If heterogeneity existed a "random effects model" was used and/or the rationale (ie, clinical appropriateness) of combining was taken into consideration (ie, was it sensible to combine), or stated explicitly.	25
Q 9.e	If homogeneity existed, the authors stated a rationale or a statistical test.	0
Q 10.a	Recognition of publication bias or file-drawer effect.	21
Q 10.b	Assessment of publication bias included graphical aids (eg, funnel plot, other available tests).	13
Q 10.c	Statistical tests (eg, Egger regression test).	0
Q 11.a	The authors provided a statement of sources of support.	79
Q 11.b	There was no conflict of interest.	50
Q 11.c	The authors provided an awareness/statement of support or conflict of interest in the primary inclusion studies.	4

A Priori Design (Q1)

All reviews included in our sample established their review design (Q1.a) and the criteria of eligibility for the selection of studies (Q1.b) before commencing with the search, collection, and data abstraction. However, most reviews suffered from a lack of clarity in framing their research questions/objectives according to the "PICO" framework (Population, Intervention, Comparison, Outcomes) recommended by methodologists and the PRISMA statement [16,52]. Although the patient population or chronic disease and the intervention under scrutiny were stated explicitly in all of the included reviews, the comparator (control) group and the outcomes of the intervention being assessed were specified in fewer cases: 25% and 67% respectively. Well-formulated research objectives addressing all 4 PICO components were identified in just 3 review articles (15%). Overall, a majority (67%) of reviews reported the patient population, the intervention, and the clinical outcomes of interest and, hence, was judged as having covered item Q1.c satisfactorily.

Duplicate Study Selection and Data Extraction (Q2)

The screening process for the selection of primary studies was performed in most cases (67%) independently, at least by 2 reviewers. Nevertheless, data extraction from the primary studies was reported as being performed independently and in duplicate in less than half of the reviews (Q2.a). In assessing the accuracy of data abstraction against primary studies in at least a sample of the included reviews as suggested by methodologists [53], we detected an instance of inappropriate coding in 1 MA [28] between the extracted data and the original publication of 1 randomized controlled trial (RCT) [54] for the outcome of congestive heart failure hospital admission. The total number of events between the control and experimental group was recorded reversely. As such, the estimated summary effect appears slightly higher and the I^2 point estimate for heterogeneity deflated (RR 0.73 [0.62-0.87] P=.0004; I^2=0 vs RR 0.78 [0.65-0.93] P=.004; I^2=46%). Data extraction was not reported being duplicated in this MA.

Out of the 24 reviews, 11 (46%) stated whether there was a consensus procedure in place or a third reviewer to resolve any disagreements (Q2.b), and 9 (38%) included a statement regarding proper resolution of existing disagreements among the reviewers (Q2.c). Overall, as shown in Multimedia Appendix 2, only one third of the reviews covered satisfactorily all of the criteria included in this domain. Additional information pertaining to the methods employed during data extraction, such as use of piloted forms/coding sheets, steps undertaken to avoid double counting of duplicate published reports, and methods used to collect additional information from the authors of the original studies were scarce.

Search Comprehensiveness (Q3)

Analysis of domain 3, which consisted of 5 criteria, showed that almost all reviews (96%) used at least 2 electronic databases to search for primary studies (Q3.a). The most prevalent databases were Medline (100%), the Cochrane Library (70%), and EMBASE (60%). All in all, 22 reviews (92%) reported the years and databases searched (Q3.b); 22 (92%) stated the keywords that were used (Q3.c); and 19 (79%) stated that the search was supplemented by reviewing the references in the studies found (Q3.d). A manual search of highly relevant journals to identify eligible studies was performed in only 3 (13%) reviews (Q3.e). Fourteen reviews (58%) used a QUOROM/PRISMA flow chart to depict and describe graphically the sequence of steps undertaken for the search and selection of relevant articles. However, presentation of the full electronic search strategy for at least 1 major database—so that one could repeat the search or assess its comprehensiveness—was made available in only 5 reviews (21%). As shown in Multimedia Appendix 2, only 2 reviews (8%) covered satisfactory all 5 criteria of the R-AMSTAR instrument within this particular domain.

Inclusion of Gray Literature (Q4)

Interestingly, most reviews focused on peer-reviewed primary studies published in English language journals. Out of the 24 articles in our database, only 2 (8%) considered the inclusion of gray literature and searched for primary studies regardless of their publication type (Q4.a). In 20 reviews (83%), the authors stated that they excluded primary studies based on their publication status (eg, abstracts, conference proceedings, and language) (Q4.b). Only one review (4%) reported that nonEnglish papers were translated (Q4.c), while 5 (21%) reported that no language restrictions were applied to the search and inclusion of studies (Q4.d).

Included and Excluded Studies Provided (Q5)

Most reviews (92%) presented a list of included studies (Q5.a), but only 25% reported a list of excluded studies in the article or in a supplement source (eg, online appendix) (Q5.b). Hence, retracing both the included and excluded studies was feasible in only 6 reviews (Q5.d). In 15 articles (63%), the authors explicitly reported the primary reasons for excluding studies (Q5.c) and subsequently reported the number of articles that were associated with each exclusion criterion. The latter item was covered satisfactory mainly by reviews that provided a PRISMA-like flow diagram [52].

Characteristics of the Included Studies (Q6)

Study-level data from the original empirical studies on the participants, interventions, and outcomes were presented in an aggregated form such as a table in 21 reviews (88%) (Q6.a). Tabulated information appeared to be complete in all of them (Q6.c). In 20 reviews (83%), the authors included in the table the ranges of the relevant PICO characteristics from the primary studies (eg, mean age of patients, duration of follow-up, severity of disease) (Q6.b).

Quality Assessment of the Primary Studies (Q7)

The methodological quality or risk of bias of the primary studies was formally appraised in 9 out of the 24 reviews (38%). In all of these, the authors provided a priori methods of assessment either in the form of a quality scale/checklist with composite scores or in the form of predefined risk of bias criteria (Q7.a). All in all, 8 reviews (33%) documented the final results of the quality appraisal in a meaningful format for each study, that is,

in the form of a grade/score or total number of criteria covered satisfactorily by each review (Q7.b). In one particular review [40], the authors stated that a risk of bias assessment was conducted according to the Cochrane Collaboration criteria, yet the results of the appraisal for each individual study were not documented. Out of the 9 reviews that assessed the quality of the primary studies, only 5 rated the level of evidence across studies or outcomes according to study design (eg, RCT, observational) and scientific quality or risk of bias of the individual studies (Q7.c). All 5 reviews (21%) used various characterized instruments to rate the overall quality of evidence (Q7.d). The most prevalent was the GRADE instrument, which was used in 3 reviews.

Tables 3 and 4 summarize the different methods, instruments, and strategies ([55-61]) used in each review to assess the quality of the included primary studies and the overall quality of the evidence. Based on the combination of these approaches, we classified the reviews under two main clusters. The first cluster focused on assessing the methodological quality of each study but did not consider the overall quality of the evidence, while the second cluster performed both assessments. Quality of evidence takes into consideration the internal validity assessment (quality or risk of bias) and design of the included studies (eg, RCT, observational), as well as other potential aspects (eg, consistency and directness of results) to rate or indicate the extent to which we can be confident that the estimated effect size or the final conclusions of the review about the effectiveness of the HT intervention are correct across each outcome of interest or individual study [55].

Table 3. Methods and instruments used for the quality assessment of the primary studies—Cluster 1.

Cluster 1	Chaudhry 2007 [26]	Clark 2007 [27]	Cox 2012 [41]	Farmer 2005 [44]
Focus of the assessment	Study design (D) Study quality (Q)	Study design (D) Study quality (Q)	Study quality (Q)	Study quality (Q)
(Focus of the assessment) Methods of assessment	(D) Inclusion of RCTs only (Q) Jüni scale [56] and York Centre criteria [57]	(D) Inclusion of RCTs only (Q) Cochrane criteria [58]	(Q) Downs and Black scale [59]	(Q) Jadad scale [60]; used only for the assessment of RCTs
Number of Assessors	NR[a]	2	2	NR
Assessors Blinded?	NR	NR	NR	NR
Adjudication or consensus procedure	NR	Yes	Yes	NR
Cross-tabulation of results for each study by domain	No	Yes	Yes	No
Overall study quality score	Yes	N/A[b]	Yes	Yes

[a]NR: not reported.

[b]N/A: non-applicable.

Table 4. Methods and instruments used for the quality assessment of the primary studies—Cluster 2.

Cluster 2	Bolton 2011 [40]	Franek 2012 [42]	Inglis 2010 [31]	Polisena 2010 [34]	MAS 2009 [46]
Focus of the Assessment	Study quality (Q) Quality of evidence (E)	Study quality (Q) Quality of evidence (E)	Study design (D) Study quality (Q) Quality of evidence (E)	Study quality (Q) Quality of evidence (E)	Study design (D) Study quality (Q) Quality of evidence (E)
(Focus of the Assessment) Methods of assessment	(Q) Cochrane criteria [58] (E) Oxford Centre for Evidence-based Medicine – Levels of Evidence	(Q) Adaptation of CONSORT statement checklist for RCTs (E) GRADE [55]	(D) Inclusion of RCTs only (Q) Cochrane criteria [58] (E) GRADE [55]	(Q) and (E) Adaptation of Hailey et al instrument [61]	(D) Inclusion of RCTs only (Q) Adaptation of the levels of evidence hierarchy proposed by Goodman (E) GRADE [55]
Number of Assessors	2	NR	2	2	NR
Assessors Blinded?	NR[a]	NR	NR	NR	NR
Adjudication or consensus procedure in place	Yes	NR	NR	NR	NR
Cross-tabulation of results for each study by domain	No	Yes	Yes	No	Yes
Overall study quality score	N/A[b]	N/A	N/A	Yes	N/A
Quality of evidence ranking	Across studies	Across outcomes	Across outcomes	Across studies	Across outcomes

[a]NR: not reported.

[b]N/A: non-applicable.

It should be noted that besides the reviews that formally appraised the quality or risk of bias of the primary studies by means of an instrument, 3 additional reviews [29,43,48] used a rating scale [62] to judge the strength of evidence of the included studies. According to this scale, the strength of evidence can be determined and appropriately ranked in 1 of 9 hierarchical levels—appearing in descending order—after considering 2 important elements: (1) the type of the design employed in each primary study (eg, large RCT, small RCT, cohort), and (2) the validity of the study based on a set of conditions of scientific rigor, including study quality. However, none of the 3 reviews conducted or considered the latter component recommended by the aforementioned scale. In the context of the analysis and formulation of conclusions, all 3 reviews ranked the evidence hierarchically according to the study design "label" of each study only. They did not critically appraise or take into consideration the actual *features* of the individual studies, which ultimately influence the risk of bias. Hence, large and small-sample RCTs were ranked higher on the hierarchy of evidence compared to nonrandomized controlled trials, cohort studies, and so on.

Scientific Quality of Included Studies Used Appropriately in Formulating Conclusions (Q8)

Out of the 9 reviews that formally assessed the scientific rigor of the primary studies (see Q7), 6 factored the results of the methodological quality into the final conclusions (Q8.a) and recommendations made for future research studies (Q8.b). Altogether, 75% of the reviews reached conclusions about the effectiveness of HT for chronic patients without considering or reflecting the potential risks of bias in the included studies. Importantly, none of the included reviews incorporated the results of the quality assessment (items in Q7) into the actual analyses of the review to explore how conclusions might be affected if studies at high risk of bias were included or excluded from the analysis.

Appropriateness of Methods Used to Combine Studies' Findings (Q9)

A majority of reviews in our database (63%) aggregated the results from the primary studies qualitatively, using narrative synthesis. However, the rationale behind the selected approach and the methods that the authors used to guide their decision were not generally mentioned. Out of 15 narrative SRs, 8 (53%) provided a statement as to why a qualitative synthesis of the evidence was chosen over a meta-analysis [26,29,33,37,40-42,48]. The primary reason in all of these reviews revolved generally around the existence of "heterogeneity" between the included studies. Nevertheless, the methods, criteria, or specific rules (eg, logic models based on the PICO framework) that were used to objectively support that a meta-analysis was not appropriate or sensible because the primary studies were clinically or methodologically too diverse, were not specified.

XSL•FO

RenderX

Out of the 15 SRs, only 3 (20%) provided an analysis plan with information about the methods, tools, or general framework that was used at each stage of the synthesis process [26,29,48]. In the remaining reviews, the logic of the decision-making process and the criteria based on which the authors assigned weights to the primary studies to arrive at final conclusions, were not specified. Moreover, the vast majority (93%) of SRs summarized and synthesized the available evidence using variants of raw data as reported in the original studies (eg, percentages, mean differences, *P* values, and counts). Only one [26] transformed the extracted data into a common statistical measure (eg, risk ratios) to allow for more transparent and direct comparisons between the observed treatment effects of the primary outcomes of interest.

As shown in Table 5, the authors of SRs used four distinct approaches to organize and synthesize the available evidence qualitatively. The most commonly used approach (in 10 SRs) was the "reported outcomes" method, in which analysis and synthesis of the results was carried out based on the most frequent outcomes assessed and reported in the original studies. Four reviews used a "levels of evidence" approach, in which the study design of the included studies was used as a basis to stratify and present the available evidence in descending order (eg, large RCTs, small RCTs, cohort studies, and case-control studies). Two of these coupled the "levels of evidence" with the "reported outcomes" method, while a third one used "vote counting" to present the direction of the intervention effect in each study (eg, positive, negative, and conflicting evidence for effect). In two of the SRs that we examined, the authors grouped and analyzed studies according to the primary mode of the telemonitoring intervention (eg, automated monitoring of signs and symptoms and telephone touch-pad-based HT modalities).

Out of the 24 reviews, 9 combined the findings from the primary studies quantitatively using meta-analytic methods. However, none of the MAs stated explicitly what criteria were used in the context of the research question(s) being addressed to support objectively that the HT trials analyzed were clinically and methodologically similar enough to be combined quantitatively (Q9.a). In one MA [34], it was stated that the quantitative pooling of study results was deemed inappropriate whenever substantial statistical heterogeneity ($I^2 \geq 50\%$) was found and this heterogeneity could not be explained by means of subgroup analysis. However, from a methodological point of view (as described later), excessive reliance on I^2 can be particularly misleading and hence, using statistical heterogeneity and point estimates of I^2 alone as the only criterion for deciding whether an MA is appropriate or not is a rather problematic strategy [63-65]. The decision to pool and present treatment estimates in an MA is not amenable to statistical tests and should be based on the clinical and methodological relevance of any heterogeneity present (eg, the age of patients, severity of disease, duration of follow-up, technology used, and study design).

As shown in Table 5, the summary statistics of the effect measures that were used in each MA were generally related to the type of investigated outcomes and available data in the original trials (ie, dichotomous, count, or continuous). The consistency of HT effects across studies was assessed and quantified for each outcome of interest in all MAs by means of a formal statistical test (Q9.b). The most common method found in 8 MAs (Table 5) involved use of the I^2 statistic, which is derived from the Chi-square test (Cochran's Q statistic). With the exemption of one [39] that reported only the range of the calculated I^2 estimates, the remaining MAs reported the precise results within the forest plots or the text of the article and provided an interpretation of the heterogeneity estimate for each investigated outcome (Q9.c).

The I^2 statistic [66,67] measures the approximate proportion of total variability in a set of treatment effect estimates that is attributable to real clinical or methodological differences between the included studies, rather than sampling error. It takes values from 0 to 100% and often thresholds (eg, 25%, 50%, and 75%) are used to make inferences about the magnitude of inconsistencies between the findings of trials [67,68]. However, simulations have shown that the I^2 statistic suffers from similar power and precision shortcomings as the Q statistic [64,65]. Thus, it can yield unreliable estimates in MAs that include a small number of trials (eg, k<15) with poor precision (ie, small number of patients and events). To this end, relevant guidelines [68] and methodologists [64,66,67,69,70] suggest that researchers should investigate, present, and consider in the interpretation of the results the 95% confidence interval (CI) of the I^2 estimate, in order to adequately reflect the uncertainty (strength of evidence) around it. That is, the spectrum of possible degrees of genuine differences between the trials in terms of treatment effects. However, none of the MAs in our database reported carrying out this statistical procedure. Although the number of included HT trials was consistently lower than 12 and most trials exhibited poor precision due to the small number of registered patients, inferences about the consistency or inconsistency of HT effects across the included trials were based on I^2 point estimates alone.

Given the potential negative implications of this methodological limitation for the reliability of MAs with respect to the interpretation of the results and choice of statistical model [70,71]), we sought to conduct a post hoc analysis to evaluate empirically the extent of uncertainty in the provided heterogeneity (I^2) estimates. As recommended [69], we used for all calculations the noncentral χ^2 based approach, which is implemented in the heterogi module of Stata (version 12.1) [72]. In total, we were able to calculate the I^2 statistic and its associated 95% CIs for all but one MA [39], for a total of 22 outcomes with 4 or more studies. Based on careful appraisal of the application and interpretation of the statistical methods used in each MA, we identified the following methodological issues.

In 6 MAs [27,28,31,34,44,47] in which the I^2 statistic was estimated to be equal to 0% for a specific outcome (Table 6), a common inference was that no heterogeneity exists or that heterogeneity is low between trials. As such, the direction and dispersion of the magnitude of clinical HT effects were interpreted as being consistent across the included trials. However, the 95% CIs, which reflect the uncertainty around these heterogeneity estimates, are particularly wide in all of

these MAs, ranging from low to high heterogeneity. As shown in Table 6, the upper limits of the 95% CI crossed into the range of large heterogeneity ($I^2 \geq 50\%$) in all of them and in 3 MAs it also exceeded or reached the 75% range (substantial heterogeneity), while the low limits of the intervals were always as low as 0%. This indicates that any strong inferences and conclusive statements about the similarity or comparability of the studies' results would be difficult to make with certainty due to the general lack of evidence. Given the poor precision of the trials included in all of these MAs, it is possible that the I^2 estimate was masked and deflated [73]. Hence, the presence of some moderate or even considerable heterogeneity between HT trials should not have been ruled out or underestimated.

The second methodological issue we identified was associated with the opposite problem, that is, overestimation of heterogeneity. In 5 forest plots of 4 MAs [28,38,46,47], in which the point estimate of I^2 was moderate (eg, 33.8%) or quite large (eg, $I^2 \geq 50\%$) (Table 6), a common inference was that there is high or even substantial inconsistency across the HT effect sizes of the trials due to genuine differences. However, as shown in Table 6, in all of these MAs the low limit of the 95% CI in the I^2 point estimates crosses into the range of little heterogeneity ($I^2 \leq 25\%$), reflecting that the evidence for large heterogeneity may not be strong enough to support the importance of the observed I^2 value. Overestimation of heterogeneity and undue reliance on I^2 estimates prompted researchers in one MA [47] to exhaust all possibilities of subgroup analysis and succumb to a poorly supported post hoc analysis in a quest for the causes of heterogeneity, while in another review it prevented the authors from carrying out an MA [34].

Last, a slightly more subtle, but yet important, methodological error concerns the issue of overweighting a study in an MA by double counting its study groups [24,74,75]. Specifically, one MA in our database [38] that compared the effects of HT with usual care on patients with hypertension, included in its sample an RCT [76] that had 1 control group (usual care with 247 patients) and 2 intervention groups: (1) blood pressure HT with Web training services (246 patients), and (2) blood pressure HT with pharmacist-assisted care via Web communications (237 patients). The way that the authors chose to handle this particular trial in their MA, for all reported outcomes, was to include it twice in each forest plot by double counting its control arm. However, the effect of this was that this particular trial was overpowered. It was counted once with 493 patients and once with 484 patients. As a result, its effective sample size appears to be 977 when in fact the true sample size was 730. This poses an important validity threat in the results of this particular review, as this trial was assigned considerable weight in all forest plots for the outcomes of interest.

With respect to the statistical model used, 6 MAs (67%) carried out random effects analyses, while 3 carried out fixed effect analyses (Q.9c). Two of latter studies [28,31] used the fixed-effect model even though some evidence of potentially moderate (eg, $I^2 > 30\%$) to substantial (eg, $I^2 > 75\%$) heterogeneity between studies was present. However, it was not justified why the fixed effect model was still deemed appropriate. In most reviews the rationale, criteria, or general assumptions that guided researchers in selecting one of two statistical models were not specified. Out of the 9 MAs, only 2 (22%) provided an explicit statement to justify the statistical model that was used to calculate the summary effects [27,31]. Both reviews were authored by the same group of researchers and focused on the effects of HT and structured telephone support (separately) versus usual care on patients with congestive heart failure. Interestingly, however, the selected model was different in each review, although the reasons or assumptions stated by the authors were almost identical.

Publication Bias (Q10)

The three criteria included in this question focus on the meta-analytic methods used to assess the likelihood of publication bias, that is, the publication or nonpublication of research findings depending on the direction of the results of the primary studies. Out of the 9 MAs included in our review, 5 considered publication bias in their assessments (Q10.a) and only 3 presented the actual funnel plots in the published article (Q10.b). In these 5 MAs, authors relied on visual inspection and interpretation of funnel plots. Formal statistical tests to assess presence of bias (eg, Egger regression test) were not used by any of the MAs (Q10.c). This is reasonable, given the small number of trials included in each review. Such tests theoretically require a considerable number of primary studies for sufficient power to detect bias; a criterion that is rarely fulfilled. However, none of the MAs acknowledged the great risk of subjectivity that is associated with visual inspection of funnel plots [70,77] and the inadequacy of this method to detect bias (let alone publication bias) when the number of studies is small (eg, k<10) or when heterogeneity is significant [78,79]. As a result, in all cases, statements about the existence of strong publication bias or absence thereof were stronger than the evidence allowed.

Conflicts of Interest (Q11)

Most reviews in our sample (79%) disclosed explicitly all the sources of support received for the conduct of the review. In 50% of them, at least one or more of the investigators were either directly affiliated or had other active involvement with entities that have competing interests in the results of the respective review, such as HT solution providers (Q10.b). Only one review (4%) examined and reported whether authors of the included empirical studies had a potential conflict of interest (Q10.c).

Table 5. Methods used in SRs and MAs to synthesize the available evidence from the primary studies.

Methods		Reviews	n
Qualitative methods (n=15)			
	Reported outcomes	[30,32,35-37,40-43,45,49]	11
	Levels of evidence (study design)	[29,32,37,48]	4[a]
	Vote counting (intervention effect)	[29]	1[a]
	Telemonitoring modality	[26,33]	2
Meta-analytic methods (n=9)			
Summary statistics			
	Risk ratios (for dichotomous data)	[27,28,31,34]	4
	Risk difference (for dichotomous data)	[27]	1[b]
	Mean difference (for continuous data)	[38,39,46]	3
	Standardized mean difference (for continuous data)	[44,47]	2
Heterogeneity			
	Assessment of heterogeneity by means of a statistical test	[27,28,31,34,38,39,44,46,47]	9
	Reported Cochran's Q statistic (Chi-square test) of heterogeneity	[27,28,31,44,46,47]	6
	Reported I^2 test of heterogeneity	[27,28,31,34,38,39,46,47]	8
Statistical model			
	Random effects meta-analysis	[27,34,38,47]	4
	Fixed effect meta-analysis	[28,31,44]	3
Meta-analysis diagnostics			
	Subgroup analysis	[34,46,47]	3
	Sensitivity analysis	[31,38]	2

[a]Includes reviews that used two different methods.

[b]Same review that used two different summary statistics.

Table 6. Confidence intervals for the I^2 estimates of MAs.

	Author (Year)	Number of trials	I^2	Low interval (95% CI)	High interval (95% CI)	Statistical model	Assessed outcomes
Heart failure							
	Clark 2007 [27]	5	0	0	79	Random effects	All-cause mortality
	Clarke 2011 [28]						
		10	51	0	76	Fixed effect	All-cause mortality
		6	59	0	83	Fixed effect	All-cause hospitalization
		6	0	0	75	Fixed effect	CHF-related hospitalization
		4	82	52	93	Fixed effect	All-cause emergency visits
	Inglis 2010 [31]						
		11	0	0	60	Fixed effect	All-cause mortality
		8	0	0	68	Fixed effect	All-cause mortality follow-up period >6 months
		8	78	56	89	Fixed effect	All-cause hospitalization
		6	85	70	93	Fixed effect	All-cause hospitalization follow-up period >6 months
		4	39	0	79	Fixed effect	CHF-related hospitalization
		4	39	0	79	Fixed effect	CHF-related hospitalization follow-up period >6 months
	Polisena 2010 [34]						
		6	0	0	75	Random effects	All-cause mortality
		4	5	0	85	Random effects	All-cause hospitalization
Hypertension							
	Omboni 2011 [38]						
		11	65.8	35	82	Random effects	Systolic blood pressure changes
		11	56.6	15	78	Random effects	Diastolic blood pressure changes
		6	77.9	44	91	Random effects	Blood pressure control
		5	79.1	50	91	Random effects	Number of antihypertensive drugs
Diabetes							
	Farmer 2005 [44]	9	0	0	65	Fixed effect	Glycemic control - Changes in HbA1c
	MAS 2009 [46]						
		7	65	20	84	Random effects	Glycemic control - Changes in HbA1c (All studies)
		4	45	0	82	Random effects	Glycemic control - Changes in HbA1c (subgroup analysis)
	Montori 2004 [47]						
		8	33.8	0	71	Random effects	Glycemic control - Changes in HbA1c
		7	0	0	71	Random effects	Glycemic control - Changes in HbA1c (post-hoc subgroup analysis)

Discussion

Principal Findings

This critical review presents the first formal and comprehensive quality assessment of published reviews that have studied the effects of HT on patients with chronic conditions. We applied the R-AMSTAR instrument to critically examine the methodological rigor and reporting characteristics of each review and also conducted a careful evaluation within the 11 domains of this particular instrument to identify risks of bias (ie, systematic errors) in inferences or results that may have affected their internal validity. To this end, R-AMSTAR was used as a general framework that guided and supported our assessment rather than a specific tool for calculating quality scores for each review. Such scores may not always reflect the true scientific quality of each review and evidence suggests that their use can be problematic in judging whether or not to trust an individual analysis, due to the potential existence of false positives or negatives [58].

The results of our bibliographic search indicate that SRs and MAs in this domain are fairly new compared to other clinical areas (eg, [18,80]). The first review was published in 2003 and focused on patients with congestive heart failure. Since then, and particularly over the last 6 years, the number of published reviews has increased substantially, while also the focus of reviewers has extended to include other chronic diseases such as chronic obstructive pulmonary disease, hypertension, and diabetes. Nonetheless, the largest body of reviews continues to focus on patients with congestive heart failure.

Based on our assessment, we found that with the recent increase in reviews of HT interventions an important number of these articles appear to lack optimal scientific rigor due to intrinsic methodological issues. Furthermore, their overall quality does not appear to have improved over time. Despite the wide availability and dissemination of important methodological guidelines [52,81] that can be utilized to guide the systematic review process and eliminate potential risks of bias, it appears that this knowledge has not yet been fully integrated in the field of HT. While several criteria were met satisfactorily by all or most reviews (eg, establishment of an a priori design (100%), reporting of inclusion/exclusion criteria (100%) and characteristics of studies (88%), use of multiple electronic searches and databases (96%)), there were other important areas that needed improvement. These areas should be considered by future SRs and MAs, in order to advance scientific progress and improve the rigor of research in the rapidly growing field of HT. As indicated by the application of the R-AMSTAR instrument and our analysis, many reviews did not perform key methodological procedures to reduce the risk of bias (eg, duplicate data extraction (42%), inclusion of gray (8%) and nonEnglish literature (21%), methodological quality assessment of included studies (38%)), and some reviews suffered from limitations in the synthesis of study results that may have affected the validity of their results and conclusions. We explain below the potential implications of these issues and provide recommendations for future reviews in this area.

Search Strategy

Although the majority of reviews used more than 2 electronic databases to search for relevant studies, other important approaches to minimize bias and enhance the search strategy were rarely used. Only 2 reviews attempted to identify primary studies in the gray literature and the vast majority restricted all searches to English articles only, although it has been demonstrated that bias can be introduced in SRs and MAs focusing exclusively on English language publications [82,83]. Inclusion and exclusion criteria established a priori for the selection of primary studies were reported explicitly in all reviews, but most failed to provide a list of references with the studies that were excluded, as recommended by methodologists and the PRISMA statement [16,52]. These methodological issues suggest a potentially limited review of the available evidence and high risk of selection and language bias. A bibliographic analysis of citation patterns that we performed confirms these concerns. Indeed, the vast majority of reviews included in our database fell short in their identification of published studies due to various languages, publication type, and date restrictions applied in the search process. The Cochrane review by Inglis et al [31], which performed the most comprehensive search among the other SRs and MAs on heart failure, provides concrete evidence of this (Multimedia Appendix 3). Concretely, these authors identified 3 relevant trials, 2 of which were published in a language other than English (one in German and one in Italian). The German publication, which was peer-reviewed, was the largest RCT (502 patients) among all trials identified by the other reviews. Nevertheless, it was not included in any of the other reviews published after 2007, neither were the other 2 RCTs that were published as abstracts, because almost all reviews restricted their search to English publications and did not consider gray literature. To minimize the risk of selection and language bias, future reviews of HT should avoid applying such restrictions as these do not align with the notion of SRs and MAs, which aim to provide a thorough and unbiased overview of all the available empirical evidence.

Discrepancies in the Inclusion of HT Studies

HT as a research area has witnessed considerable growth over the past decade. Nevertheless, from a conceptual point of view there seems to be a lack of consensus between authors of SRs and MAs in the terminology they used (eg, "telecare" [47], "telemedicine" [44], "telehealth" [41], "telehealth and remote monitoring" [29]), and most importantly in the types of interventions and technologies that qualify as HT. For instance, Chaudhry et al [26] argue in their review that there is no clear rational for excluding telephone-based interventions that use one-on-one telephone calls between nurses and patients, while other reviewers contend the opposite (eg, [31,32,48,49]. The protocol of our critical appraisal and in particular the examination of citation patterns revealed several discordant views between the included reviews on the inclusion, classification, and analysis of certain interventions. The majority of reviews strongly converged on the inclusion of interventions that were based on telemetry devices offering automated or message-based monitoring and transmission of physiologic signs or symptoms through communication networks (see

Multimedia Appendix 3). However, there were important disagreements between reviews in the inclusion and analysis of other interventions such as stand-alone telephone support (52, 63), automated telephone calls, toll-free computerized voice answering systems (13, 31), videophone (70), television-based support (4), video-conferencing (46), and website-based support (35).

The following example provides a good illustration of the problem that currently exists and the consequences it has on the results and direct comparison of the results of HT reviews. An RCT that was included in 3 reviews of HT for heart failure [28,29,31], 2 MAs and 1 SR, comprised a control group of usual care and 2 intervention groups. The first intervention group was assigned to structured telephone support, while the second was assigned to videophone that did not involve any automated monitoring or transmission of vital signs and symptoms. The 2 reviews [28,29] considered the videophone intervention as home telemonitoring, while the third one did not [31]. The way the third review chose to treat this study was to combine both intervention groups into one and analyze them quantitatively as structured telephone support. This indicates that there is no commonly agreed upon definition of HT and its core properties. Future research should address this important issue by proposing and validating a taxonomy that would capture the different types/forms of HT and enable robust comparisons across trials.

Quality Assessment of Included Studies

The validity of the results produced by prior reviews and the confidence in their conclusions depend to a large extent on the quality of the included studies. There is ample evidence showing that the scientific quality of primary studies is not always adequate and methodological flaws, when not identified and accounted for, may inflate or deflate the results of an SR [84-86]. Current guidelines [52,58] suggest two different quality assessments that must be performed by reviewers in each review: the methodological quality (or risk of bias) of the original studies and the quality of evidence [55] to indicate the extent to which we can be confident that an estimate of effect or the final conclusions of a review are correct across each outcome of interest. There also exist various strategies [58] that may be applied to incorporate the results of these assessments in the analysis and conclusions of the review. Unfortunately, our findings within the particular area of HT are rather disappointing and raise important concerns. Out of the 24 reviews, only 9 (38%) assessed the methodological quality of the included studies and 5 of them (21%) rated the overall quality of the available evidence. Furthermore, only 4 reviews factored the results of the quality assessment in their final conclusions. Therefore, the possibility that biased studies have inflated or deflated the results of prior reviews of HT cannot be ruled out.

Selection and Justification of the Data Synthesis Method

Decisions concerning the selection of the data synthesis method that is most appropriate for addressing the research question(s) of the review require thoughtful consideration, as well as clinical judgment and should be based on explicit clinical and methodological criteria that minimize subjectivity as much as possible [68]. Based on the results of our evaluation, the

rationale and criteria used to guide and support the decision of the researchers to synthesize the available evidence narratively or quantitatively was not always evident. Out of the 15 qualitative SRs, 8 (53%) provided some explanation for not conducting an MA, but even in these cases the criteria used to decide that studies were not clinically or methodologically similar enough to be pooled were not revealed. On their part, MAs of HT did not provide a rationale or a statement specifying what criteria were used to support the decision to combine statistically studies that may vary in terms of patients' stages of severity, home telemonitoring approaches, implementation settings, and other important aspects. This finding indicates that most reviewers may use narrative synthesis or meta-analysis as a "default action", based on methodological preferences or prior experiences rather than explicit and clinically relevant criteria that minimize subjectivity. However, it would be informative for future reviews to address this issue by clearly specifying any methods or specific rules (eg, logic models based on the PICO framework) that were used to guide the selection of a particular synthesis approach [6,22].

Qualitative Synthesis of Studies

Authors employing narrative or qualitative synthesis should describe explicitly the analysis plan underpinning each stage of the evidence synthesis process, in order to clarify and support the logic that was used to reach the final conclusions. Presenting an analysis plan is of paramount importance and should be an integral part of the Methods section in future SRs of HT, as it clarifies the synthesis process, improves the transparency and reliability of the review, and acts as a safeguard against bias that can arise from placing inappropriate emphasis on the results of one study over another [87-89]. Such an analysis plan must incorporate among others appropriate techniques for the transformation of raw data to a common statistical or numerical measure (eg, risk ratios, mean differences) across studies selected for inclusion [87]. This will allow reviewers to develop meaningful summaries of effect sizes that can facilitate robust and transparent comparisons across the range of studied effects. Unfortunately, the majority of narrative SRs failed to meet these criteria and in most cases review authors tended to rely excessively on reported P values, which have a notorious record for being misleading, particularly in situations with small primary studies that have large within-study variance (ie, poor precision) and are not sufficiently powered to reach significant results [74]. Given the inherent risks of misinterpreting nonsignificant results as evidence of no effect, future SRs in this area should preferably synthesize the available data by estimating effect-sizes from each primary study (as it was done in one of the SRs [26]) rather than reported P values.

Measuring Inconsistency of HT Effects in Meta-Analyses

One of the main objectives of the statistical methods used in MAs of HT interventions is to evaluate the dispersion among the results of the included studies, that is, the between-study heterogeneity in effect sizes, in order to assess the consistency of study findings. In light of observed heterogeneity, it is important to investigate and explain, whenever possible, what is causing it in order to increase scientific understanding and

clinical relevance. With respect to the first goal, all 9 MAs included in our sample adhered closely to recommended guidelines and assessed formally the variability (heterogeneity) of the HT studies' results by calculating either Cochran's Q, I^2, or both heterogeneity statistics in most cases. This was particularly encouraging and reflects a good practice that is generally consistent with other MAs in the health care domain [17]. However, the limitations of these metrics [64-67,70,71] and the uncertainty around the I^2 point estimates, which can be expressed with 95% confidence intervals, were not considered in any of the included MAs. As a result, firm claims or inferences about the extent of inconsistencies in the HT effects between trials in most cases were stronger than the evidence allowed. Perhaps this limitation can be attributed to the fact that the Review Manager (RevMan) software, which was used in more than half of the MAs, does not provide users with a functionality to calculate the confidence intervals of I^2. This is an issue that has also been highlighted by other researchers and communicated in hope that future updates of this software will make confidence intervals an integral part of I^2 heterogeneity calculations [73].

Future MAs in this area should continue to use both statistics to measure the statistical significance and proportion of heterogeneity in the observed effects. However, the limitations of these metrics must be taken into consideration. The Q statistic is subject to the same caveats as all tests of significance and should always be interpreted with due caution based on the number of HT studies included in the analysis [70]. The I^2 is not precise and hence, confidence intervals for I^2 estimates should always be reported and interpreted carefully, as they are valuable for reflecting the uncertainty associated with the estimated ratio of true heterogeneity to total variation in the observed effects [69]. When the number of primary studies included in an MA is limited (eg, k<15) and the within-study variance is large, the I^2 estimate should be interpreted with caution and any strong statements about the consistency of the observed HT effects "should be avoided or tempered appropriately, regardless of the results" [70]. Furthermore, when the sizes of HT effects vary substantially, as was the case with certain outcomes in some MAs (eg, [31,38]), this variance in the results should become the primary focus in the discussion of an MA and the summary effect should be less important or even not important at all [74].

The Choice Between Fixed and Random Effects Meta-Analysis

When combining data from various HT studies, a major dilemma is to decide whether to perform a fixed or random effects meta-analysis. This decision is particularly important as the choice of model might affect the estimate of the effect size and, ultimately, the interpretation of the results [79,90]. A fixed effect MA of HT interventions is based on the premise that all studies included in the review are functionally identical and are estimating a common (fixed) treatment effect [74,91]. That is, there are no genuine differences; all factors that potentially could influence the observed effect size such as the nature of the intervention (eg, sophistication of the technology, frequency

of data transmission, home visits, and educational support) are functionally the same in all studies. Thus, any observed between-study variation (ie, statistical heterogeneity) in the results is attributed only to sampling error. On the other hand, random effects MA is based on the premise that the observed estimates of treatment effect are not identical in the included HT studies but follow some distribution. That is, they vary from study to study because of genuine differences (eg, in the nature of the intervention) as well as sampling variability (chance). Studies may differ in the mix of participants (eg, stages of severity), the quality, or implementation of the intervention, and so on. Hence, each study is estimating a different underlying effect. As such, a fixed effect MA provides an estimate of a "common" treatment effect, while the summary result produced by random effects MA provides an estimate of the "average" treatment effect [74,90]. It is also important to note that from a statistical point of view, when the between-study variance (statistical heterogeneity) is 0%, random effects analysis is reduced and coincides with a fixed effect analysis, showing similar effects anyhow. However, in the presence of any between-study heterogeneity, fixed effect meta-analyses provide overly precise summary results with narrower confidence intervals than random effects meta-analyses [90]. As we present next, this can lead to spuriously lower levels of statistical significance for the summary effects and may wrongly imply that a "common" treatment effect exists when in reality there are real differences in treatment effects across studies [79,90].

Our evaluation revealed that the random effects model, which facilitates a broader outlook as it summarizes the distribution of the intervention effects across studies, appears to be the most preferable statistical model among MAs of HT interventions. Indeed, from a clinical perspective, the "one size fits all" approach of the fixed effect model appears to be difficult to justify. The participants and contextual characteristics of HT interventions in most cases differ in many practical ways that may have an impact on the results [22]. It is implausible that effect modifiers in HT studies such as the technology, patients, program characteristics, and risks of bias are functionally identical or equivalent across all the included trials. Both HT and usual care have evolved dramatically over the past 15 years and these temporal changes may have affected the results of the included trials, resulting in greater heterogeneity. Nevertheless, 2 MAs on heart failure [28,31] applied the fixed effect model, despite the functional differences between the trials and the presence of moderate (eg, I^2>30%) to substantial (eg, I^2>75%) statistical heterogeneity in the observed effects. The use of a fixed rather than a random effects model influenced their results, as it produced tighter confidence intervals and spuriously low levels of statistical significance for the effects of HT. Specifically, in the Cochrane review the effect estimate for all cause-hospitalization using the fixed effect model showed a statistically significant (P=.02) reduction of 9% favoring HT (RR 0.91, 95% CI 0.84-0.99). Whereas the random effects model yields a nonsignificant (P=.22) effect size of the same magnitude with a wider confidence interval (RR 0.91, 95% CI 0.78-1.06), reflecting the uncertainty behind the positive effects of HT on average. Similarly, in the MA by Clarke et al [28], the effect estimate for mortality using the fixed effect model shows a

significant (*P*=.02) reduction of deaths by 23% in favor of HT (RR 0.77, 95% CI 0.61-0.97). However, the random effects model yields a more conservative and nonsignificant (*P*=.30) effect-size of 17% on average with wider confidence interval (RR 0.83, 95% CI 0.58-1.19), reflecting again that the underlying effect of HT may not always be positive across all patients and contexts. Given the clinical and methodological differences of the HT trials included in these 2 MAs, the use of the fixed effect model appears to be counterintuitive and the a priori assumptions that led to its selection should have been revisited, especially after the detection of statistical heterogeneity [74]. Future MAs of HT interventions should comply with methodological guidelines and describe explicitly the rationale and the criteria that were used to choose between fixed and random effects meta-analysis. Also, when the random effects approach is used, then the pooled results should be interpreted appropriately as the "average" effect of the HT intervention [90,91], as was done in one of the MAs [34] in our sample.

Limitations

If we apply the critical review approach to our own review, we realize that a number of challenges were faced in the process of appraising the methodological quality of the included SRs and MAs, which may have in turn affected our findings. First, our appraisal was performed on the basis of the information reported, explicitly or implicitly, in each review. Therefore, as in all methodological quality or risk of bias assessments, the accuracy of the judgments made by the evaluators relies heavily on the reporting adequacy of the reviews. It is possible that the authors conducted their review more rigorously. However, being aware of the length restrictions imposed by the journals and in light of competing demands for reporting the main findings of their review, they might have decided to omit some methodological information that was perceived as subtle or less important to report. It is also possible that the peer-review process itself resulted in abbreviating the text to meet space limitations. One recommendation for future reviews to alleviate this issue is to provide essential details about the protocol of the review in an electronic version, as is the practice in several peer-reviewed journals today, to aid in understanding the systematic review process considered. On the other hand, peer-reviewed journals that have an interest in publishing SRs and MAs in the area of HT should devote space for publishing online supplementary material and adopt appropriate mechanisms for flagging problems with and allowing corrections of previous work, once errors or other important deficiencies have been identified [24]. Also, the research community must be prepared to validate the results of reviews, in order to correct them if necessary and the results must be published in such a way that will facilitate this process [24]. We conducted a post hoc analysis and found that out of the 16 journals in which the included reviews were published, 10 (63%) allowed the publication of online appendices but only 3 reviews provided an appendix or a supplement file.

Second, it is important to note that the findings of our evaluation are confined to the reviews that met our inclusion criteria described in the Methods section. Although our bibliographic search identified several "narrative reviews" that focus on the effectiveness of HT interventions on patients with various chronic diseases, when these were not self-identified as systematic or did not feature essential properties of an SR or MA, they were excluded from our study. This strict selection process may have contributed to an overestimation of the methodological quality of HT reviews as reflected by the R-AMSTAR instrument and our analysis. Also excluded were several reviews that provided an all-inclusive and mixed overview of HT interventions along with various other "remote monitoring" interventions (eg, structured telephone support and stand-alone video consultation), but did not make a clear distinction between them in the analysis of the results. Therefore, our findings are not generalizable to reviews in which HT was one among many other multidisciplinary interventions of remote patient monitoring, although most would agree that the highlighted methodological deficiencies have significant relevance and are applicable to these reviews as well.

Conclusion

This study is the first attempt to evaluate the overall quality of prior SRs and MAs of HT interventions. The comprehensiveness of the search strategy used to identify relevant reviews, the duplicated process in relation to study selection, data extraction, and quality appraisal, as well as the use of a validated instrument that offers the ability to conduct an in-depth quality assessment, are key indicators of the methodological soundness of the present study.

The number of published SRs and MAs in the area of HT has substantially increased in the last decade offering to a wide range of health care stakeholders an extensive base of "large-scale evidence" from the synthesis of multiple primary studies on the clinical, behavioral, structural, and economic effects of HT for patients with chronic conditions. Yet, despite the significant body of knowledge that has been developed, wide acceptance by payers and care providers and integration of HT as an effective patient management approach remains problematic. This is mainly because the existing knowledge base still exhibits several important methodological weaknesses and research gaps.

Of utmost importance, our critical assessment revealed that the overall quality and rigor of existing SRs and MAs of HT interventions is highly variable, with no signs of improvement over time. An important number of reviews contain several common methodological shortcomings that impair their internal validity and limit their usefulness for clinical, educational, research, and policy purposes. As a result, a range of questions regarding the effectiveness of HT for chronic disease management remain unanswered, including which is the ideal and most effective combination of case management and remote monitoring, which behavior change techniques and modalities are most effective, whether the effectiveness of interventions is influenced by participant demographics and settings, and whether HT is an effective and viable solution from an economic point of view. We thus recommend that future reviews in this area improve their overall rigor as well as their reporting aspects by adhering closely to available methodological guidelines. More precisely, they should at least include the following elements: (1) clearly stated research question(s) explicitly

describing the patient population, intervention, comparison intervention, and outcomes; (2) comprehensive and clearly stated search strategies; (3) formal appraisal of the validity of the primary studies (ie, risk of bias assessment) with appropriate attempts to explore the impact of studies with high risk of bias on the estimated effects of HT; and (4) more rigorous methods of data synthesis with transparent descriptions and justifications of the techniques or statistics used.

To conclude, it is our hope that this study will contribute to increase the overall quality of SRs and MAs in the HT area, as well as in the broader telehealth domain, by helping authors minimize diverse risks of biases and avoid previous methodological deficiencies. Nonetheless, we believe that building more rigorous and stronger evidence in the HT area will require unprecedented efforts by researchers, clinicians, funders, journal editors, and peer reviewers. Such efforts include but are not limited to the involvement of individuals with both clinical and methodological expertise in the conduct of SRs and MAs; amendments to the general instructions published by the journals with specific guidelines or links to methodological and reporting recommendations; the involvement of individuals in the peer-review process with prior experience and knowledge in the methodologies of SRs and MAs; and adoption of mechanisms to allow updates or corrections of online published material to address important deficiencies or even errors identified after publication.

Conflicts of Interest

None declared.

Multimedia Appendix 1

List of excluded articles.

[PDF File (Adobe PDF File), 29KB - jmir_v15i7e150_app1.pdf]

Multimedia Appendix 2

R-AMSTAR assessment of methodological rigor for each review.

[PDF File (Adobe PDF File), 409KB - jmir_v15i7e150_app2.pdf]

Multimedia Appendix 3

Citation analysis of systematic reviews and meta-analyses of home telemonitoring interventions for patients with heart failure.

[PDF File (Adobe PDF File), 627KB - jmir_v15i7e150_app3.pdf]

References

1.   World Health Organization. Global status report on noncommunicable diseases 2010: Description of the global burden of NCDs, their risk factors and determinants. Geneva, Switzerland: World Health Organization; 2011:1-176.
2.   Mathers CD, Loncar D. Projections of global mortality and burden of disease from 2002 to 2030. PLoS Med 2006 Nov;3(11):e442 [FREE Full text] [doi: 10.1371/journal.pmed.0030442] [Medline: 17132052]
3.   Shaw JE, Sicree RA, Zimmet PZ. Global estimates of the prevalence of diabetes for 2010 and 2030. Diabetes Res Clin Pract 2010 Jan;87(1):4-14. [doi: 10.1016/j.diabres.2009.10.007] [Medline: 19896746]
4.   Meystre S. The current state of telemonitoring: a comment on the literature. Telemed J E Health 2005 Feb;11(1):63-69. [doi: 10.1089/tmj.2005.11.63] [Medline: 15785222]
5.   Paré G, Poba-Nzaou P, Sicotte C. Home telemonitoring for chronic disease management: an economic assessment. Int J Technol Assess Health Care 2013 Apr;29(2):155-161. [doi: 10.1017/S0266462313000111] [Medline: 23514722]
6.   Anker SD, Koehler F, Abraham WT. Telemedicine and remote management of patients with heart failure. Lancet 2011 Aug 20;378(9792):731-739. [doi: 10.1016/S0140-6736(11)61229-4] [Medline: 21856487]
7.   Roine R, Ohinmaa A, Hailey D. Assessing telemedicine: a systematic review of the literature. CMAJ 2001 Sep 18;165(6):765-771 [FREE Full text] [Medline: 11584564]
8.   Chien E, Geffen M, Gordon D. A Study Commissioned by Canada Health Infoway. Telehealth Benefits and Adoption-Connecting People and Providers Across Canada URL: https://www.infoway-inforoute.ca/index.php/component/docman/doc_download/188-avantages-et-adoption-de-la-telesante-rapport-complet [accessed 2013-07-11] [WebCite Cache ID 6I280zh4K]
9.   Stroetmann KA, Artmann J, Stroetmann VN. Final European progress report: European Commission—Directorate General Information Society and Media, ICT for Health Unit. European countries on their journey towards national eHealth

infrastructures URL: http://www.ehealth-strategies.eu/report/eHealth_Strategies_Final_Report_Web.pdf [accessed 2013-07-11] [WebCite Cache ID 6I28v9qh1]

10. Darkins A, Ryan P, Kobb R, Foster L, Edmonson E, Wakefield B, et al. Care Coordination/Home Telehealth: the systematic implementation of health informatics, home telehealth, and disease management to support the care of veteran patients with chronic conditions. Telemed J E Health 2008 Dec;14(10):1118-1126. [doi: 10.1089/tmj.2008.0021] [Medline: 19119835]

11. Cartwright M, Hirani SP, Rixon L, Beynon M, Doll H, Bower P, Whole Systems Demonstrator Evaluation Team. Effect of telehealth on quality of life and psychological outcomes over 12 months (Whole Systems Demonstrator telehealth questionnaire study): nested study of patient reported outcomes in a pragmatic, cluster randomised controlled trial. BMJ 2013;346:f653 [FREE Full text] [Medline: 23444424]

12. Palmer M, Steffen C, Iakovidis I, Giorgio F. European Commission perspective: Telemedicine for the benefit of patients. Chronic disease management and remote patient monitoring. Health Care Systems and Society, Eurohealth 2009;15(1):13-15.

13. European Commission. On telemedicine for the benefit of patients, healthcare systems and society. 2008. URL: http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=COM:2008:0689:FIN:EN:PDF [accessed 2013-07-11] [WebCite Cache ID 6I28mwSND]

14. Cook DJ, Mulrow CD, Haynes RB. Systematic reviews: synthesis of best evidence for clinical decisions. Ann Intern Med 1997 Mar 1;126(5):376-380. [Medline: 9054282]

15. Mulrow CD. Rationale for systematic reviews. BMJ 1994 Sep 3;309(6954):597-599 [FREE Full text] [Medline: 8086953]

16. Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gøtzsche PC, Ioannidis JP, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. PLoS Med 2009 Jul 21;6(7):e1000100 [FREE Full text] [doi: 10.1371/journal.pmed.1000100] [Medline: 19621070]

17. Moher D, Tetzlaff J, Tricco AC, Sampson M, Altman DG. Epidemiology and reporting characteristics of systematic reviews. PLoS Med 2007 Mar 27;4(3):e78 [FREE Full text] [doi: 10.1371/journal.pmed.0040078] [Medline: 17388659]

18. Jadad AR, Moher M, Browman GP, Booker L, Sigouin C, Fuentes M, et al. Systematic reviews and meta-analyses on treatment of asthma: critical evaluation. BMJ 2000 Feb 26;320(7234):537-540 [FREE Full text] [Medline: 10688558]

19. Papageorgiou SN, Papadopoulos MA, Athanasiou AE. Evaluation of methodology and quality characteristics of systematic reviews in orthodontics. Orthod Craniofac Res 2011 Aug;14(3):116-137. [doi: 10.1111/j.1601-6343.2011.01522.x] [Medline: 21771267]

20. Kelly KD, Travers A, Dorgan M, Slater L, Rowe BH. Evaluating the quality of systematic reviews in the emergency medicine literature. Ann Emerg Med 2001 Nov;38(5):518-526. [doi: 10.1067/mem.2001.115881] [Medline: 11679863]

21. MacDonald SL, Canfield SE, Fesperman SF, Dahm P. Assessment of the methodological quality of systematic reviews published in the urological literature from 1998 to 2008. J Urol 2010 Aug;184(2):648-653. [doi: 10.1016/j.juro.2010.03.127] [Medline: 20639030]

22. Gurné O, Conraads V, Missault L, Mullens W, Vachierys JL, Van Mieghem W, Belgian Working Group on Heart Failure. A critical review on telemonitoring in heart failure. Acta Cardiol 2012 Aug;67(4):439-444. [Medline: 22997998]

23. Casas JP, Kwong J, Ebrahim S. Telemonitoring for chronic heart failure: not ready for prime time. Cochrane Database Syst Rev 2011;2011:ED000008. [doi: 10.1002/14651858.ED000008] [Medline: 21833931]

24. Senn SJ. Overstating the evidence: double counting in meta-analysis and related problems. BMC Med Res Methodol 2009;9:10 [FREE Full text] [doi: 10.1186/1471-2288-9-10] [Medline: 19216779]

25. Green S, Higgins JPT, Alderson P, Clarke M, Mulrow CD, Oxman AD. Introduction. In: Higgins JP, Green S, editors. Cochrane Handbook for Systematic Reviews of Interventions. New Jersey: John Wiley & Sons, Ltd; 2008:1-9.

26. Chaudhry SI, Phillips CO, Stewart SS, Riegel B, Mattera JA, Jerant AF, et al. Telemonitoring for patients with chronic heart failure: a systematic review. J Card Fail 2007 Feb;13(1):56-62 [FREE Full text] [doi: 10.1016/j.cardfail.2006.09.001] [Medline: 17339004]

27. Clark RA, Inglis SC, McAlister FA, Cleland JG, Stewart S. Telemonitoring or structured telephone support programmes for patients with chronic heart failure: systematic review and meta-analysis. BMJ 2007 May 5;334(7600):942 [FREE Full text] [doi: 10.1136/bmj.39156.536968.55] [Medline: 17426062]

28. Clarke M, Shah A, Sharma U. Systematic review of studies on telemonitoring of patients with congestive heart failure: a meta-analysis. J Telemed Telecare 2011;17(1):7-14. [doi: 10.1258/jtt.2010.100113] [Medline: 21097564]

29. Dang S, Dimmick S, Kelkar G. Evaluating the evidence base for the use of home telehealth remote monitoring in elderly with heart failure. Telemed J E Health 2009 Oct;15(8):783-796. [doi: 10.1089/tmj.2009.0028] [Medline: 19831704]

30. Giamouzis G, Mastrogiannis D, Koutrakis K, Karayannis G, Parisis C, Rountas C, et al. Telemonitoring in chronic heart failure: a systematic review. Cardiol Res Pract 2012;2012:410820 [FREE Full text] [doi: 10.1155/2012/410820] [Medline: 22720184]

31. Inglis SC, Clark RA, McAlister FA, Ball J, Lewinter C, Cullington D, et al. Structured telephone support or telemonitoring programmes for patients with chronic heart failure. Cochrane Database Syst Rev 2010(8):CD007228. [doi: 10.1002/14651858.CD007228.pub2] [Medline: 20687083]

32. Louis AA, Turner T, Gretton M, Baksh A, Cleland JGF. A systematic review of telemonitoring for the management of heart failure. Eur J Heart Fail 2003 Oct;5(5):583-590. [Medline: 14607195]

33. Maric B, Kaan A, Ignaszewski A, Lear SA. A systematic review of telemonitoring technologies in heart failure. Eur J Heart Fail 2009 May;11(5):506-517. [doi: 10.1093/eurjhf/hfp036] [Medline: 19332417]

34. Polisena J, Tran K, Cimon K, Hutton B, McGill S, Palmer K, et al. Home telemonitoring for congestive heart failure: a systematic review and meta-analysis. J Telemed Telecare 2010;16(2):68-76. [doi: 10.1258/jtt.2009.090406] [Medline: 20008054]

35. Seto E. Cost comparison between telemonitoring and usual care of heart failure: a systematic review. Telemed J E Health 2008 Sep;14(7):679-686. [doi: 10.1089/tmj.2007.0114] [Medline: 18817497]

36. AbuDagga A, Resnick HE, Alwan M. Impact of blood pressure telemonitoring on hypertension outcomes: a literature review. Telemed J E Health 2010 Sep;16(7):830-838. [doi: 10.1089/tmj.2010.0015] [Medline: 20815751]

37. Jaana M, Paré G, Sicotte C. Hypertension Home Telemonitoring: Current Evidence and Recommendations for Future Studies. Disease Management & Health Outcomes 2007;15(1):19-31.

38. Omboni S, Guarda A. Impact of home blood pressure telemonitoring and blood pressure control: a meta-analysis of randomized controlled studies. Am J Hypertens 2011 Sep;24(9):989-998. [doi: 10.1038/ajh.2011.100] [Medline: 21654858]

39. Verberk WJ, Kessels AG, Thien T. Telecare is a valuable tool for hypertension management, a systematic review and meta-analysis. Blood Press Monit 2011 Jun;16(3):149-155. [doi: 10.1097/MBP.0b013e328346e092] [Medline: 21527847]

40. Bolton CE, Waters CS, Peirce S, Elwyn G, EPSRC and MRC Grand Challenge Team. Insufficient evidence of benefit: a systematic review of home telemonitoring for COPD. J Eval Clin Pract 2011 Dec;17(6):1216-1222. [doi: 10.1111/j.1365-2753.2010.01536.x] [Medline: 20846317]

41. Cox NS, Alison JA, Rasekaba T, Holland AE. Telehealth in cystic fibrosis: a systematic review. J Telemed Telecare 2012 Mar;18(2):72-78. [doi: 10.1258/jtt.2011.110705] [Medline: 22198961]

42. Franek J. Home telehealth for patients with chronic obstructive pulmonary disease (COPD): an evidence-based analysis. Ont Health Technol Assess Ser 2012;12(11):1-58 [FREE Full text] [Medline: 23074421]

43. Jaana M, Paré G, Sicotte C. Home telemonitoring for respiratory conditions: a systematic review. Am J Manag Care 2009 May;15(5):313-320 [FREE Full text] [Medline: 19435399]

44. Farmer A, Gibson OJ, Tarassenko L, Neil A. A systematic review of telemedicine interventions to support blood glucose self-monitoring in diabetes. Diabet Med 2005 Oct;22(10):1372-1378. [doi: 10.1111/j.1464-5491.2005.01627.x] [Medline: 16176199]

45. Jaana M, Paré G. Home telemonitoring of patients with diabetes: a systematic assessment of observed effects. J Eval Clin Pract 2007 Apr;13(2):242-253. [doi: 10.1111/j.1365-2753.2006.00686.x] [Medline: 17378871]

46. Medical Advisory Secretariat. Home telemonitoring for type 2 diabetes: an evidence-based analysis. Ont Health Technol Assess Ser 2009;9(24):1-38 [FREE Full text] [Medline: 23074529]

47. Montori VM, Helgemoe PK, Guyatt GH, Dean DS, Leung TW, Smith SA, et al. Telecare for patients with type 1 diabetes and inadequate glycemic control: a randomized controlled trial and meta-analysis. Diabetes Care 2004 May;27(5):1088-1094. [Medline: 15111526]

48. Paré G, Moqadem K, Pineau G, St-Hilaire C. Clinical effects of home telemonitoring in the context of diabetes, asthma, heart failure and hypertension: a systematic review. J Med Internet Res 2010;12(2):e21 [FREE Full text] [doi: 10.2196/jmir.1357] [Medline: 20554500]

49. 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]

50. Inglis SC, Clark RA, McAlister FA, Stewart S, Cleland JG. Which components of heart failure programmes are effective? A systematic review and meta-analysis of the outcomes of structured telephone support or telemonitoring as the primary component of chronic heart failure management in 8323 patients: Abridged Cochrane Review. Eur J Heart Fail 2011 Sep;13(9):1028-1040. [doi: 10.1093/eurjhf/hfr039] [Medline: 21733889]

51. Kung J, Chiappelli F, Cajulis OO, Avezova R, Kossan G, Chew L, et al. From Systematic Reviews to Clinical Recommendations for Evidence-Based Health Care: Validation of Revised Assessment of Multiple Systematic Reviews (R-AMSTAR) for Grading of Clinical Relevance. Open Dent J 2010;4:84-91 [FREE Full text] [Medline: 21088686]

52. Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. PLoS Med 2009 Jul 21;6(7):e1000097 [FREE Full text] [doi: 10.1371/journal.pmed.1000097] [Medline: 19621072]

53. Whitlock EP, Lin JS, Chou R, Shekelle P, Robinson KA. Using existing systematic reviews in complex systematic reviews. Ann Intern Med 2008 May 20;148(10):776-782. [doi: 10.7326/0003-4819-148-10-200805200-00010] [Medline: 18490690]

54. Dar O, Riley J, Chapman C, Dubrey SW, Morris S, Rosen SD, et al. A randomized trial of home telemonitoring in a typical elderly heart failure population in North West London: results of the Home-HF study. Eur J Heart Fail 2009 Mar;11(3):319-325 [FREE Full text] [doi: 10.1093/eurjhf/hfn050] [Medline: 19174529]

55. Atkins D, Best D, Briss PA, Eccles M, Falck-Ytter Y, Flottorp S, GRADE Working Group. Grading quality of evidence and strength of recommendations. BMJ 2004 Jun 19;328(7454):1490 [FREE Full text] [doi: 10.1136/bmj.328.7454.1490] [Medline: 15205295]

56. Jüni P, Altman DG, Egger M. Systematic reviews in health care: Assessing the quality of controlled clinical trials. BMJ 2001 Jul 7;323(7303):42-46 [FREE Full text] [doi: 10.1136/bmj.323.7303.42] [Medline: 11440947]

XSL•FO
RenderX

57.   Centre for Reviews and Dissemination. Undertaking systematic review of research on effectiveness. York: University of York; 2001.

58.   Higgins JPT, Altman DG. Assessing Risk of Bias in Included Studies. In: Higgins JP, Green S, editors. Cochrane Handbook for Systematic Reviews of Interventions. New Jersey: John Wiley & Sons, Ltd; 2008:187-241.

59.   Downs SH, Black N. The feasibility of creating a checklist for the assessment of the methodological quality both of randomised and non-randomised studies of health care interventions. J Epidemiol Community Health 1998 Jun;52(6):377-384 [FREE Full text] [Medline: 9764259]

60.   Jadad AR, Moore RA, Carroll D, Jenkinson C, Reynolds DJ, Gavaghan DJ, et al. Assessing the quality of reports of randomized clinical trials: Is blinding necessary? Controlled Clinical Trials 1996 Feb;17(1):1-12. [doi: 10.1016/0197-2456(95)00134-4] [Medline: 8721797]

61.   Hailey D, Ohinmaa A, Roine R. Study quality and evidence of benefit in recent assessments of telemedicine. J Telemed Telecare 2004;10(6):318-324. [doi: 10.1258/1357633042602053] [Medline: 15603628]

62.   Jovell AJ, Navarro-Rubio MD. Evaluation of scientific evidence. Med Clin (Barc) 1995 Dec 2;105(19):740-743. [Medline: 8523956]

63.   Ioannidis JP, Patsopoulos NA, Rothstein HR. Reasons or excuses for avoiding meta-analysis in forest plots. BMJ 2008 Jun 21;336(7658):1413-1415 [FREE Full text] [doi: 10.1136/bmj.a117] [Medline: 18566080]

64.   Huedo-Medina TB, Sánchez-Meca J, Marín-Martínez F, Botella J. Assessing heterogeneity in meta-analysis: Q statistic or I2 index? Psychol Methods 2006 Jun;11(2):193-206. [doi: 10.1037/1082-989X.11.2.193] [Medline: 16784338]

65.   Rücker G, Schwarzer G, Carpenter JR, Schumacher M. Undue reliance on I(2) in assessing heterogeneity may mislead. BMC Med Res Methodol 2008;8:79 [FREE Full text] [doi: 10.1186/1471-2288-8-79] [Medline: 19036172]

66.   Higgins JP, Thompson SG. Quantifying heterogeneity in a meta-analysis. Stat Med 2002 Jun 15;21(11):1539-1558. [doi: 10.1002/sim.1186] [Medline: 12111919]

67.   Higgins JP, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. BMJ 2003 Sep 6;327(7414):557-560 [FREE Full text] [doi: 10.1136/bmj.327.7414.557] [Medline: 12958120]

68.   Deeks JJ, Higgins J, Altman DG. Analysing Data and Undertaking Meta‐Analyses. In: Higgins JP, Green S, editors. Cochrane Handbook for Systematic Reviews of Interventions. New Jersey: John Wiley & Sons, Ltd; 2008:243-296.

69.   Ioannidis JP, Patsopoulos NA, Evangelou E. Uncertainty in heterogeneity estimates in meta-analyses. BMJ 2007 Nov 3;335(7626):914-916 [FREE Full text] [doi: 10.1136/bmj.39343.408449.80] [Medline: 17974687]

70.   Ioannidis JP. Interpretation of tests of heterogeneity and bias in meta-analysis. J Eval Clin Pract 2008 Oct;14(5):951-957. [doi: 10.1111/j.1365-2753.2008.00986.x] [Medline: 19018930]

71.   Ioannidis JP, Patsopoulos NA, Evangelou E. Uncertainty in heterogeneity estimates in meta-analyses. BMJ 2007 Nov 3;335(7626):914-916 [FREE Full text] [doi: 10.1136/bmj.39343.408449.80] [Medline: 17974687]

72.   HETEROGI: Stata module to quantify heterogeneity in meta-analysis [computer program]. Version revised 25 Jan 2006: Statistical Software Components S449201, Boston College Department of Economics; 2005. URL: http://econpapers. repec.org/software/bocbocode/s449201.htm [accessed 2013-07-16] [WebCite Cache ID 6IA5RAmkQ]

73.   Thorlund K, Imberger G, Johnston BC, Walsh M, Awad T, Thabane L, et al. Evolution of heterogeneity (I2) estimates and their 95% confidence intervals in large meta-analyses. PLoS One 2012;7(7):e39471 [FREE Full text] [doi: 10.1371/journal.pone.0039471] [Medline: 22848355]

74.   Borenstein M, Hedges L, Higgins J, Rothstein H. Introduction to Meta-Analysis. Hoboken, NJ: John Wiley & Sons Inc; 2009.

75.   Ortiz de Guinea A, Webster J, Staples DS. A meta-analysis of the consequences of virtualness on team functioning. Information & Management 2012 Oct;49(6):301-308. [doi: 10.1016/j.im.2012.08.003]

76.   Green BB, Cook AJ, Ralston JD, Fishman PA, Catz SL, Carlson J, et al. Effectiveness of home blood pressure monitoring, Web communication, and pharmacist care on hypertension control: a randomized controlled trial. JAMA 2008 Jun 25;299(24):2857-2867 [FREE Full text] [doi: 10.1001/jama.299.24.2857] [Medline: 18577730]

77.   Terrin N, Schmid CH, Lau J. In an empirical evaluation of the funnel plot, researchers could not visually identify publication bias. J Clin Epidemiol 2005 Sep;58(9):894-901. [doi: 10.1016/j.jclinepi.2005.01.006] [Medline: 16085192]

78.   Lau J, Ioannidis JP, Terrin N, Schmid CH, Olkin I. The case of the misleading funnel plot. BMJ 2006 Sep 16;333(7568):597-600 [FREE Full text] [doi: 10.1136/bmj.333.7568.597] [Medline: 16974018]

79.   Kavvoura FK, Ioannidis JP. Methods for meta-analysis in genetic association studies: a review of their potential and pitfalls. Hum Genet 2008 Feb;123(1):1-14. [doi: 10.1007/s00439-007-0445-9] [Medline: 18026754]

80.   Choi PT, Halpern SH, Malik N, Jadad AR, Tramèr MR, Walder B. Examining the evidence in anesthesia literature: a critical appraisal of systematic reviews. Anesth Analg 2001 Mar;92(3):700-709 [FREE Full text] [Medline: 11226105]

81.   The Cochrane Collaboration. In: Higgins JP, Green S, editors. Cochrane Handbook for Systematic Reviews of Interventions. Chichester, UK: John Wiley & Sons, Ltd; 2008.

82.   Egger M, Zellweger-Zähner T, Schneider M, Junker C, Lengeler C, Antes G. Language bias in randomised controlled trials published in English and German. Lancet 1997 Aug 2;350(9074):326-329. [doi: 10.1016/S0140-6736(97)02419-7] [Medline: 9251637]

83. Grégoire G, Derderian F, Le Lorier J. Selecting the language of the publications included in a meta-analysis: Is there a tower of babel bias? Journal of Clinical Epidemiology 1995 Jan;48(1):159-163. [doi: 10.1016/0895-4356(94)00098-B] [Medline: 7853041]

84. Egger M, Juni P, Bartlett C, Holenstein F, Sterne J. How important are comprehensive literature searches and the assessment of trial quality in systematic reviews? Empirical study. Health Technol Assess 2003;7(1):1-76 [FREE Full text] [Medline: 12583822]

85. Moher D, Pham B, Jones A, Cook DJ, Jadad AR, Moher M, et al. Does quality of reports of randomised trials affect estimates of intervention efficacy reported in meta-analyses? Lancet 1998 Aug 22;352(9128):609-613. [doi: 10.1016/S0140-6736(98)01085-X] [Medline: 9746022]

86. Schulz KF, Chalmers I, Hayes RJ, Altman DG. Empirical evidence of bias. Dimensions of methodological quality associated with estimates of treatment effects in controlled trials. JAMA 1995 Feb 1;273(5):408-412. [Medline: 7823387]

87. Popay J, Roberts H, Sowden A, Petticrew M, Arai L, Rodgers M, et al. Guidance on the conduct of narrative synthesis in systematic reviews. A product from the ESRC methods programme Version 1. 2006. URL: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.178.3100&rep=rep1&type=pdf

88. Rodgers M, Sowden A, Petticrew M, Arai L, Roberts H, Britten N, et al. Testing Methodological Guidance on the Conduct of Narrative Synthesis in Systematic Reviews: Effectiveness of Interventions to Promote Smoke Alarm Ownership and Function. Evaluation 2009 Jan 01;15(1):49-73. [doi: 10.1177/1356389008097871]

89. Arai L, Britten N, Popay J, Roberts H, Petticrew M, Rodgers M, et al. Testing methodological developments in the conduct of narrative synthesis: a demonstration review of research on the implementation of smoke alarm interventions. Evidence & Policy: A Journal of Research, Debate and Practice 2007 Aug 01;3(3):361-383. [doi: 10.1332/174426407781738029]

90. Riley RD, Higgins JP, Deeks JJ. Interpretation of random effects meta-analyses. BMJ 2011;342:d549. [Medline: 21310794]

91. Riley RD, Gates S, Neilson J, Alfirevic Z. Statistical methods can be improved within Cochrane pregnancy and childbirth reviews. J Clin Epidemiol 2011 Jun;64(6):608-618. [doi: 10.1016/j.jclinepi.2010.08.002] [Medline: 21109399]

Abbreviations

GRADE: grading of recommendations assessment, development, and evaluation
HT: home telemonitoring
MA: meta-analysis
PICO: population, intervention, comparison, outcomes
PRISMA: preferred reporting items for systematic reviews and meta-analyses
QUOROM: quality of reporting of meta-analyses
R-AMSTAR: revised assessment of multiple systematic reviews
RCT: randomized controlled trial
SR: systematic review

XSL•FO
RenderX

<u>Original Paper</u>

Facilitating Out-of-Home Caregiving Through Health Information Technology: Survey of Informal Caregivers' Current Practices, Interests, and Perceived Barriers

Donna M Zulman[1,2], MD, MS; John D Piette[3,4], PhD; Emily C Jenchura[1], BS; Steven M Asch[1,2], MD, MPH; Ann-Marie Rosland[3,4], MD, MS

[1]Center for Health Care Evaluation, VA Palo Alto Health Care System, Menlo Park, CA, United States

[2]Division of General Medical Disciplines, Stanford University, Stanford, CA, United States

[3]Center for Clinical Management Research, VA Ann Arbor, Ann Arbor, MI, United States

[4]Division of General Internal Medicine, University of Michigan Medical School, Ann Arbor, MI, United States

Corresponding Author:
Donna M Zulman, MD, MS
Center for Health Care Evaluation
VA Palo Alto Health Care System
MC-152
795 Willow Road
Menlo Park, CA, 94025
United States
Phone: 1 650 493 5000 ext 29113
Fax: 1 650 617 2736
Email: donna.zulman@va.gov

Abstract

Background: Many patients with chronic conditions are supported by out-of-home informal caregivers—family members, friends, and other individuals who provide care and support without pay—who, if armed with effective consumer health information technology, could inexpensively facilitate their care.

Objective: We sought to understand caregivers' use of, interest in, and perceived barriers to health information technology for out-of-home caregiving.

Methods: We conducted 2 sequential Web-based surveys with a national sample of individuals who provide out-of-home caregiving to an adult family member or friend with a chronic illness. We queried respondents about their use of health information technology for out-of-home caregiving and used multivariable regression to investigate caregiver and care-recipient characteristics associated with caregivers' technology use for caregiving.

Results: Among 316 out-of-home caregiver respondents, 34.5% (109/316) reported using health information technology for caregiving activities. The likelihood of a caregiver using technology increased significantly with intensity of caregiving (as measured by number of out-of-home caregiving activities). Compared with very low intensity caregivers, the adjusted odds ratio (OR) of technology use was 1.88 (95% CI 1.01-3.50) for low intensity caregivers, 2.39 (95% CI 1.11-5.15) for moderate intensity caregivers, and 3.70 (95% CI 1.62-8.45) for high intensity caregivers. Over 70% (149/207) of technology nonusers reported interest in using technology in the future to support caregiving. The most commonly cited barriers to technology use for caregiving were health system privacy rules that restrict access to care-recipients' health information and lack of familiarity with programs or websites that facilitate out-of-home caregiving.

Conclusions: Health information technology use for out-of-home caregiving is common, especially among individuals who provide more intense caregiving. Health care systems can address the mismatch between caregivers' interest in and use of technology by modifying privacy policies that impede information exchange.

(J Med Internet Res 2013;15(7):e123) doi:10.2196/jmir.2472

XSL•FO
RenderX

Introduction

As American society becomes increasingly mobile and households decline in size [1], many chronically ill patients find themselves without a caregiver in their home. By some accounts, as many as 70% of the estimated 66 million Americans who provide unpaid assistance to ill or older adults live apart from their care recipient [2,3]. Out-of-home caregiving is especially common when the care recipient is an elderly relative [4], and this trend is likely to become more pronounced as the population ages [5].

Out-of-home caregivers face a number of unique challenges. Some caregivers experience emotional stress, guilt, and helplessness related to living apart from loved ones who are in need of care [6-8]. Additional challenges may arise when caregivers live at an increased distance from their care recipients [3,9]. In one assessment, costs associated with out-of-home caregiving doubled as travel increased from 1 to 3 hours (US $386 per month) to more than 3 hours (US $674 per month). Long-distance caregivers reported spending 22 hours per month on average assisting their care recipients with instrumental activities, such as transportation, shopping, and finances, and more than half of them visited their care recipient at least a few times per month, despite a mean travel distance of 450 miles [9].

Health information technology may provide an opportunity to support out-of-home caregivers' activities. Consumer health information technology encompasses a wide range of technologies that allow patients to participate in their health care via electronic means [10]. Examples of health information technology include electronic personal health records, applications that facilitate chronic condition management (eg, programs for tracking blood pressure and glucose), Internet resources with medication and disease information, and tools that facilitate communication with health care providers. Many of these applications may be of value to out-of-home caregivers as well, for example by alleviating uncertainty about a care recipient's symptoms and status, or enhancing information exchange with a care recipient's health care team.

Despite recent discussions that health information technology could facilitate caregiving from afar [4], there have been few assessments of current and potential technology use for this purpose. We conducted a survey of individuals who care for an out-of-home adult family member or friend with a chronic condition. Our objectives were to (1) determine rates of, and interest in, health information technology use for out-of-home caregiving activities, (2) examine caregiver and patient characteristics associated with technology use, and (3) identify barriers to out-of-home caregivers' use of technology that may be overcome through enhanced technology and associated policies.
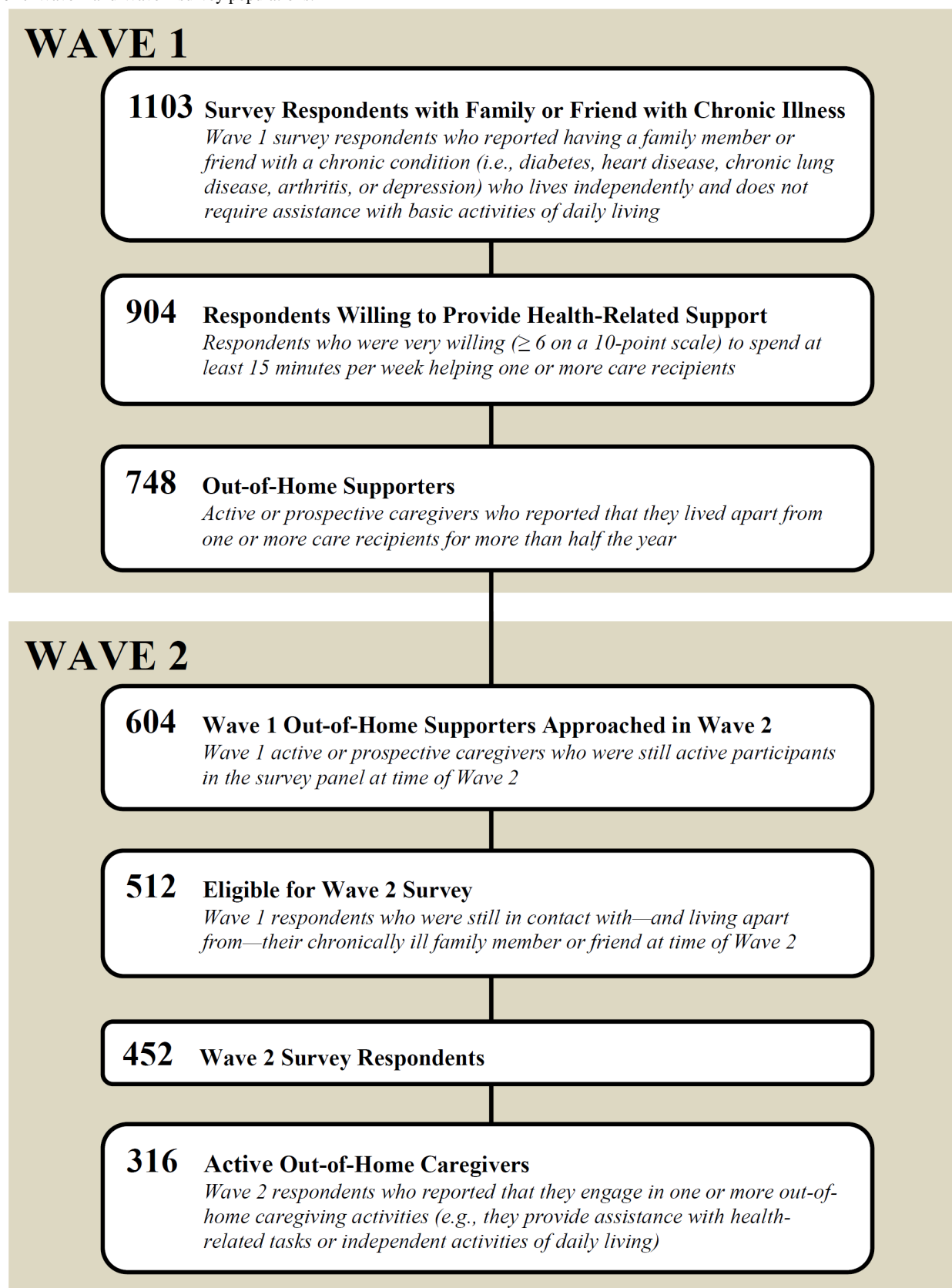
Methods

Survey Design and Administration

This paper reports findings from 2 sequential surveys of individuals who support family members and friends with chronic illness. The surveys for this study were administered by Knowledge Networks, a research firm that maintains a large, nationally representative survey panel of adults. Knowledge Networks' panel is very similar to the United States population with respect to race/ethnicity, gender, educational attainment, and income [11]. In return for their participation in the panel, members receive Internet access and computing equipment at no cost [12,13].

For this study, we identified potential participants using data from a previous study (Wave 1: January 26 to February 16, 2010) [14]. We identified 748 individuals from the Wave 1 cohort who (1) had an adult family member or friend with a chronic illness (including diabetes, chronic heart disease, chronic lung disease, arthritis, and/or depression), (2) lived apart from this person for more than half of the year, and (3) reported a high willingness to help this person with his or her health. Of note, individuals whose care recipients were living in a long-term care facility or required assistance with basic activities of daily living were excluded from the Wave 1 cohort because the focus of this earlier study was on support for independent and ambulatory adults with a chronic illness [14].

For the current study (Wave 2: January 20 to February 21, 2011), we invited the 604 Wave 1 participants who were still active Knowledge Networks panelists to complete a follow-up survey about their use of health information technology to support their out-of-home care recipient (Multimedia Appendix 1). Of the 512 individuals who completed a screening questionnaire (response rate 84.8%), 452 reported that they were still in touch with—and living apart from—the care recipient whom they had identified in Wave 1. In this paper, we report survey findings for the subgroup of 316 respondents who we identified as active out-of-home caregivers (ie, they reported engaging in 1 or more out-of-home caregiving activities, described subsequently, to support their care recipient) (Figure 1). Specifically, we report these respondents' use of, interest in, and barriers to health information technology for out-of-home caregiving activities, and we describe caregiver and patient characteristics associated with technology use.

Figure 1. Wave 1 and Wave 2 survey populations.

WAVE 1

1103 **Survey Respondents with Family or Friend with Chronic Illness**
Wave 1 survey respondents who reported having a family member or friend with a chronic condition (i.e., diabetes, heart disease, chronic lung disease, arthritis, or depression) who lives independently and does not require assistance with basic activities of daily living

904 **Respondents Willing to Provide Health-Related Support**
Respondents who were very willing (≥ 6 on a 10-point scale) to spend at least 15 minutes per week helping one or more care recipients

748 **Out-of-Home Supporters**
Active or prospective caregivers who reported that they lived apart from one or more care recipients for more than half the year

WAVE 2

604 **Wave 1 Out-of-Home Supporters Approached in Wave 2**
Wave 1 active or prospective caregivers who were still active participants in the survey panel at time of Wave 2

512 **Eligible for Wave 2 Survey**
Wave 1 respondents who were still in contact with—and living apart from—their chronically ill family member or friend at time of Wave 2

452 **Wave 2 Survey Respondents**

316 **Active Out-of-Home Caregivers**
Wave 2 respondents who reported that they engage in one or more out-of-home caregiving activities (e.g., they provide assistance with health-related tasks or independent activities of daily living)

XSL•FO
RenderX

Survey Measures

Dependent Variable: Use of Health Information Technology for Out-of-Home Caregiving

Our dependent variable was use of health information technology for out-of-home caregiving activities. We asked Wave 2 survey respondents, "In the past year, in what ways have you used the computer, Internet, or email to help [your care recipient] manage his or her health." Textbox 1 indicates the response options provided to respondents. We dichotomized respondents based on whether they reported any versus no use of 1 or more of these technologies for caregiving.

Textbox 1. Survey questions regarding use of health information technology for caregiving.

In the past year, in what ways have you used the computer, Internet, or email to help [your care recipient] manage his or her health:

1. I helped him or her find health information online

2. I sent messages to his or her doctor or other health care provider by email

3. I helped him or her track his or her health information (for example, their blood pressure, blood sugar, or medication use) on a computer

4. I helped him or her access his or her health records through a system linked to his or her health care provider (ie, a personal health record system or health portal)

5. I helped him or her use a health portal or personal health record system that is available through his or her health care provider

6. I helped him or her fill medications or medical supplies online

7. I helped him or her look up medical test results online

8. Other ways: _____

9. I have not used the computer, Internet, or email for any of the above

---

Independent Variables: Caregiver and Care Recipient Characteristics

We measured out-of-home caregiving intensity by assessing respondents' involvement in health-related activities that might be amenable to out-of-home assistance. These included assistance with independent activities of daily living in the past 3 months, assistance with health-related tasks in the past 3 months, discussions about health with the care recipient usually or always in conversations, communication with the care recipient's physician in the past year, and guidance given to the care recipient about questions to ask a health care provider in the past year. After examining the distribution of respondents' participation in these activities, we generated an out-of-home caregiving intensity index comprising the sum of these items (1 = very low, 2 = low, 3 = moderate, 4-5 = high).

We queried caregivers about their comfort with technology (including computers, the Internet, email, text messaging, and learning how to use new programs on a computer or the Internet), using survey questions that were previously piloted in an evaluation of a Web-based caregiving intervention [15]. Factor analysis demonstrated that all of the questions loaded onto a single factor (Cronbach alpha=.89); thus, the 5 items were combined into a single measure comprising the sum of technology modalities or tasks with which respondents agreed or strongly agreed they felt comfortable (0 = very low, 1-2 = low, 3-4 = moderate, 5 = high).

We obtained additional information about caregiver characteristics from the Wave 1 survey and from the Knowledge Networks database of panel members, including caregivers' age, sex, race/ethnicity, household income, education, health status, and whether the respondent had Internet access prior to enrolling in Knowledge Networks.

Finally, we obtained caregiver-reported information about care recipients' characteristics. In Wave 1, we asked caregivers to rank their care recipient's health status (5-point scale, poor to excellent) [16] and to report whether their care recipient had a hospital admission or emergency room visit in the past year. In addition, we constructed a single variable, unmet health or health care needs, based on whether the care recipient sometimes to frequently discussed any of the following issues with the caregiver over the past 6 months: pain or bothersome symptoms, medication side effects, confusion about a doctor's advice, unanswered questions that were asked of the doctor, or insufficient support to manage his or her health problems. We also collected information about the care recipient's relationship with the caregiver and their geographic distance from one another (Wave 1), and about the care recipient's age and whether he or she uses the Internet (Wave 2).

Additional Descriptive Variables

If respondents indicated that they did not have experience using technology for a specific caregiving purpose, we asked about their interest in using technology for that purpose in the future if it would help their care recipients improve their health. We also queried all respondents about barriers to technology use for out-of-home caregiving, including insufficient time, unfamiliarity with relevant programs or websites, health problems, and privacy rules that restrict access to their care recipients' health information.

Data Analysis

We first examined rates of health information technology use and interest in technology for specific caregiving activities. We then used multivariable logistic regression models to examine the association between out-of-home caregiving intensity and a respondent's use of technology for caregiving activities, adjusting for comfort with technology, as well as caregiver's age, sex, race/ethnicity, education, income, and health status.

We constructed a similar model to investigate whether specific care recipient characteristics were associated with a caregiver's use of technology. For this model, we included the care recipient's age, health status, incidence of hospitalizations, and incidence of emergency room visits over the previous year, the presence of unmet health or health care needs, the care recipient's geographic distance from the caregiver, and whether the care recipient uses the Internet. Finally, we examined common barriers cited by technology-using and technology-nonusing caregivers that prevent them from using technology (or using it more frequently) for caregiving activities.

We used Stata 12.0 (StataCorp LP, College Station, TX, USA) to perform all analyses. Rates of item-level missing data were less than 8% for all covariates used in analyses. Regression diagnostic procedures yielded no evidence of multicollinearity in any of the regression models. Datasets were deidentified before receipt from Knowledge Networks. Both waves of the study were classified as exempt by the Institutional Review Board at the University of Michigan.

Results

Characteristics of the 316 survey respondents and their care recipients are described in Table 1. There were 109 (34.5%) out-of-home caregivers who reported using health information technology for caregiving activities, 24 (26.1%) of whom reported a frequency of monthly or more. Among these technology users, the most common purpose for technology use was to help a care recipient find health information online (70.6%), whereas sending emails to health care providers, tracking health information, accessing health records, filling medications, and looking at medical test results online were each cited by fewer than 15% of technology users.

Among the 207 respondents who reported no use of technology for caregiving, 122 (58.9%) stated that the reason for this was that their care recipient did not need their help in this way. However, 150 (73.0%) expressed a willingness to use technology in the future if it would help their care recipient with his or her health, for example 139 (67.8%) to find health information, 111 (53.6%) to track personal health information, and 104 (50.2%) to fill medications or medical supplies (Table 2). In addition, 90 of the 109 (83.0%) active technology users were interested in expanding their technology use in the future to support at least 1 additional caregiving task that they were

not already engaging in using technology. Of note, active technology users were interested in expanding their technology use to interact with their care recipients' health care system, for example to communicate with health care providers (57/101, 56.4%) and help their care recipients look up medical test results online (68/102, 66.7%).

Multivariable logistic regression revealed that greater out-of-home caregiving intensity was significantly associated with caregivers' likelihood of using health information technology. Compared to respondents with very low intensity caregiving roles, the adjusted odds of caregiving-related technology use increased steadily when caregiving intensity was low (adjusted odds ratio [OR] 1.88, 95% CI 1.01-3.50, P=.05), moderate (adjusted OR 2.39, 95% CI 1.11-5.15, P=.03), and high (adjusted OR 3.70, 95% CI 1.62-8.45, P=.002) (Table 3). The likelihood of technology use also increased markedly with a caregiver's comfort using technology. None of the other caregiver characteristics that we assessed were associated with technology use (Multimedia Appendix 2).

In a separate multivariable logistic regression analysis investigating whether care recipient characteristics (including age, geographic distance from the caregiver, and health status) were associated with a respondent's use of technology for caregiving, no significant relationships were observed (Multimedia Appendix 3).

Nearly half (49.4%, 156/316) of all respondents, 40.1% (83/207) of technology nonusers, and 67.0% (73/109) of technology users reported that there were barriers to their use of technology for out-of-home caregiving. Among the respondents reporting barriers to technology use, 57.7% (28.5% of all respondents) cited privacy rules of their care recipient's health care provider, and 58.3% (28.8% of all respondents) cited unfamiliarity with programs or websites that facilitate out-of-home caregiving. In contrast, very few respondents reported that insufficient time, computer/Internet complexity, distrust in the Internet, or their own health limitations impeded their use of technology for caregiving (Table 4). There were few differences in the frequency of barriers cited by technology users and technology nonusers, with the exception that active technology users were more than twice as likely as nonusers to report that privacy rules impeded their use of technology for caregiving (53/109, 48.6% vs 37/207, 17.9%, respectively).

Table 1. Description of study population (N=316).[a]

Characteristics	n (%)
Out-of-home caregivers	
Age	
18-29	38 (12.0)
30-44	86 (27.2)
45-59	101 (32.0)
≥60	91 (28.8)
Female	199 (63.0)
Education	
Less than high school	29 (9.2)
High school degree	72 (22.8)
Some college	112 (35.4)
College degree or higher	103 (32.6)
Race/ethnicity	
White, Non-Hispanic	189 (59.8)
Black, Non-Hispanic	71 (22.5)
Hispanic	56 (17.7)
Geographic Region	
Northwest	57 (18.0)
Midwest	62 (19.6)
South	130 (41.1)
West	67 (21.2)
Out-of-home caregiving activities (time frame)	
Assistance with independent activities of daily living (past 3 months) (N=314)	138 (44.0)
Assistance with health-related tasks (past 3 months) (N=312)	69 (22.1)
Frequent discussions about health with care recipient (N=316)	131 (41.5)
Phone conversations with care recipient's doctor (past 12 months) (N=310)	43 (13.9)
Suggested questions for care recipient to ask health care provider (past 12 months) (N=308)	262 (85.1)
Independent Internet access[b]	241 (76.3)
Comfort with technology	
Very low/low	88 (27.9)
Moderate	114 (36.1)
High	114 (36.1)
Care recipients [c]	
Age (N=313)	
<50	68 (21.7)
60-64	90 (28.8)
65-74	83 (26.5)
≥75	72 (23.0)
Use Internet	183 (57.9)

Characteristics	n (%)
Health status	
Very good or excellent	53 (16.8)
Good	121 (38.3)
Fair or poor	142 (44.9)
Relationship with caregiver	
Spouse/partner	4 (1.3)
Adult child	26 (8.2)
Sibling	88 (27.9)
Parent or parent-in-law	124 (39.2)
Other relative/friend	74 (23.4)
Distance from caregiver (N=312)	
<5 miles	74 (23.7)
5-20	85 (27.2)
21-100	44 (14.1)
>100	109 (34.9)

[a]N=316 unless otherwise specified.

[b]Knowledge Networks provides Internet access to panel participants who do not have independent access.

[c]All care recipient characteristics are caregiver-reported.

Table 2. Health information technology functions that are of interest to out-of-home caregivers for adults with chronic conditions.

Technology function	Current technology nonusers, % (proportion of respondents)	Current technology users[a], % (proportion of respondents)
Help care recipient find health information online	67.8 (139/205)	80.7 (25/31)
Help care recipient track his or her health information (eg, blood pressure, blood sugar, or medications)	53.6 (111/207)	61.6 (61/99)
Help care recipient look up medical test results online	52.2 (108/207)	66.7 (68/102)
Help care recipient use a health portal or personal health record system	51.2 (106/207)	56.7 (59/104)
Help care recipient fill medications or medical supplies online	50.2 (104/207)	54.3 (51/94)
Help care recipient keep track of his or her health records on the computer	49.3 (101/205)	63.1 (65/103)
Send email messages to care recipient's doctor or other health care provider	44.9 (92/205)	56.4 (57/101)
Interest in one or more of the above functions	73.0 (150/207)	83.0 (90/109)

[a]Respondents who reported current technology use for one or more caregiving tasks were asked about their interest in expanding their use of technology for additional caregiving tasks in the future if it would help their care recipient manage his or her health.

Table 3. Out-of-home caregivers' characteristics associated with their use of health information technology to support individuals with chronic conditions[a] (N=316; 301 of whom are in multivariate model).

Caregiver characteristics	n (%)	Unadjusted OR (95% CI)	Adjusted OR (95% CI)
Out-of-home caregiving intensity [b]			
Very low	125 (39.6)		
Low	104 (32.9)	1.82 (1.03-3.23)	1.88 (1.01-3.50)
Moderate	49 (15.5)	2.18 (1.08-4.41)	2.39 (1.11-5.15)
High	38 (12.0)	3.91 (1.83-8.36)	3.70 (1.62-8.45)
Caregiver's comfort with technology [b]			
Very low	46 (14.6)		
Low	42 (13.3)	1.64 (0.61-4.42)	1.23 (0.41-3.67)
Moderate	114 (36.1)	2.31 (1.01-5.26)	2.09 (0.87-5.02)
High	114 (36.1)	2.88 (1.27-6.54)	3.49 (1.34-9.11)

[a]Multivariable logistic regression model adjusted for caregiver's age, education, income, race/ethnicity, and health status (see Multimedia Appendix 2 for results from full model).

[b]Categories described in detail in Multimedia Appendix 2. In the presented analysis, caregiving intensity was analyzed as categorical indicator variables. When caregiving intensity was analyzed as a continuous variable in a secondary analysis, the relationship with technology use had an adjusted OR of 1.54 (95% CI 1.20-1.98, P=.001).

Table 4. Barriers to health information technology use for out-of-home caregiving.[a]

Barriers	Current technology nonusers, % (n=207)	Current technology users, % (n=109)
Unfamiliarity with programs or websites that facilitate out-of-home caregiving	29.5	27.5
Privacy rules of care recipient's health care provider	17.9	48.6
Insufficient time	7.3	12.8
Computer/Internet too complicated	8.2	5.5
Distrust in Internet for health-related information	6.8	3.7
Health or functional limitations	1.5	1.8
One or more of the above barriers	40.1	67.0

[a]Health information technology nonusers and users were asked to indicate all of the barriers that impede their use of technology or their more frequent use of technology, respectively, to help their care recipients with their health.

Discussion

Principal Findings

In this national survey of out-of-home caregivers for a chronically ill family member or friend, more than one-third (34.5%, 109/316) reported using health information technology to facilitate caregiving activities, and technology use was significantly more common among caregivers providing more intensive support. Interest in technology for caregiving far exceeded active use, suggesting an opportunity for technology innovation and expansion to better meet the needs of these individuals and their care recipients. Our findings also highlight important information-sharing barriers that can be addressed by health systems to more fully engage out-of-home caregivers in the health care of chronically ill patients.

According to a recent Pew Internet survey, close to 80% of caregivers now have access to the Internet, and approximately two-thirds of online caregivers report that their last Internet

health information search was on behalf of another person, suggesting that use of technology to support informal caregiving activities is pervasive [17]. Few studies, however, have investigated technology use and its desirability among caregivers who live apart from their care recipients. A recent National Alliance for Caregiving report revealed that individuals providing care from a distance were more likely than their in-home counterparts to report that technology could make them more effective as caregivers [18]. Our study builds on this report by describing specific technology applications that are used most frequently by out-of-home caregivers, and by identifying barriers to technology use among these individuals.

One technological feature of great interest to out-of-home caregivers in our study (both active technology users and technology nonusers) is the ability to interact with their care recipient's health care system, for example to communicate with a provider or monitor laboratory results. Many of these out-of-home caregiving technology functions could potentially be performed through a patient portal or electronic personal

XSL•FO
RenderX

health record (PHR) [19,20]—tools that are increasingly available through various health care systems, including Kaiser and the Veterans Health Administration. Several studies have documented growing interest in adapting PHRs to enhance information sharing among patients, their caregivers, and their network of health care providers. For example, in a study of more than 18,000 users of the Veterans Affairs' My Health*e*Vet PHR, approximately 80% expressed interest in sharing access to their record with a family member, caregiver, or provider outside the Veterans Affairs system [21]. Other studies indicate that caregivers are similarly interested in having remote access to their care recipient's electronic health information [22,23].

Unfortunately, despite patient and caregiver preferences for information sharing, many health care systems impose barriers that limit such communication. In our survey, nearly half of technology-using caregivers (48.6%, 53/109) indicated that health system privacy rules impede their ability to use technology for out-of-home caregiving activities. Patients who wish to share their electronic health information are frequently limited in terms of the specific individuals to whom they may authorize access, and the process is often cumbersome and may require legal documentation [24-26]. Although these regulations stem from reasonable data security considerations, when too restrictive, they may prevent patients from using PHR systems in the ways they find most valuable [27]. Our findings suggest that health systems should consider delegation applications that enable patients to easily share their electronic health information with caregivers.

An additional barrier to technology use for caregiving was unfamiliarity with available programs, despite the fact that Web-based and mobile applications designed specifically for caregivers abound [28,29]. A previous survey of caregivers (both in-home and out-of-home) identified other obstacles to caregivers' use of technology, including perceived cost (37%) and potential resistance by the care recipient (20%) [18]. These findings suggest that current technologies are either not adequately disseminated to or are not meeting the needs of caregivers and their care recipients, and that the implementation of existing caregiving technology would benefit from a greater user-centered focus.

It should be noted that we used a broad definition of caregiving for this study, including all individuals who engage in at least 1 of 5 common out-of-home caregiving activities. Historically, the term *caregiver* has been used to refer to individuals who provide fairly intense and task-oriented care [9], but there is growing awareness that many caregivers do much more than assist with basic activities of daily living. Caregivers commonly help with chronic illness management tasks, such as medication adherence, tracking of blood pressure or sugar, communication with patients' health care providers, and health system navigation [30,31]. Because these tasks are not reliant on

physical proximity, they may be particularly amenable to support through technology.

Limitations

Several limitations to our findings warrant discussion. First, although Knowledge Networks maintains a nationally representative panel, the subset of participants who met our criteria might not represent all out-of-home caregivers for adults with chronic illness. In addition, because all of Knowledge Networks' panelists have Internet access, either independently or as compensation for their panel participation, rates of technology use for caregiving may be higher among survey respondents than in the general population (where 78% have Internet access) [32]. Second, the asynchronous nature of our surveys may have resulted in certain characteristics of survey respondents (eg, caregiving intensity) and care recipients (eg, health status) changing between Wave 1 and Wave 2. Third, we relied exclusively on self-reported data, which could have resulted in recall bias, especially with regards to caregiving intensity and care recipients' health care utilization. Fourth, our assessment of technology use for out-of-home caregiving may be an underestimate because (1) we queried survey respondents about their use of technology to care for only 1 out-of-home care recipient even if they provide care to multiple individuals, and (2) our survey did not include some emerging caregiving technologies, such as telehealth, videoconferencing, and mobile applications. Finally, this study focused on caregiving for chronically ill adults who are independent in basic activities of daily living; thus, findings cannot be generalized to caregivers of children or individuals with severe cognitive or functional impairments, such as dementia.

Conclusions

In conclusion, our study suggests that health information technology use is common among out-of-home caregivers for adults with chronic conditions, especially among those providing more intensive care. Both active users and nonusers of technology indicated high levels of interest in expanding their use of technology and adopting new applications for caregiving purposes. The gap between interest and use, as well as barriers cited by survey respondents, should guide technology development and regulations to better address the needs of out-of-home caregivers. Additional investigation is needed to further elucidate specific technology features that are of greatest value to out-of-home caregivers and their care recipients, and to identify the applications that most improve chronic disease management and clinical outcomes. Out-of-home caregivers, armed with remote access to patient health information and their health care team, represent a promising opportunity to enhance chronic disease care, although we need to develop thoughtful implementation procedures and policy to ensure that we achieve this potential.

XSL•FO
RenderX

Conflicts of Interest

None declared.

Multimedia Appendix 1

Survey about technology use among out-of-home caregivers.

[PDF File (Adobe PDF File), 14KB - jmir_v15i7e123_app1.pdf]

Multimedia Appendix 2

Full multivariable logistic regression model of caregiver characteristics associated with health information technology use for out-of-home caregiving (N=301).

[PDF File (Adobe PDF File), 197KB - jmir_v15i7e123_app2.pdf]

Multimedia Appendix 3

Multivariable logistic regression model of care recipient characteristics associated with caregiver's use of health information technology to facilitate out-of-home caregiving (N=302).

[PDF File (Adobe PDF File), 79KB - jmir_v15i7e123_app3.pdf]

References

1. US Census Bureau. 2004 Oct 15. Average population per household and family: 1940 to present URL: http://www.census.gov/population/socdemo/hh-fam/tabHH-6.pdf [accessed 2012-11-26] [WebCite Cache ID 6CT972DaT]
2. Donelan K, Hill CA, Hoffman C, Scoles K, Feldman PH, Levine C, et al. Challenged to care: Informal caregivers in a changing health system. Health Aff (Millwood) 2002;21(4):222-231 [FREE Full text] [Medline: 12117133]
3. National Alliance for Caregiving and the AARP. Caregiving in the U.S. 2009 URL: http://www.caregiving.org/data/Caregiving_in_the_US_2009_full_report.pdf [accessed 2012-11-25] [WebCite Cache ID 6CT9ibYnK]
4. Benefield LE, Beck C. Reducing the distance in distance-caregiving by technology innovation. Clin Interv Aging 2007;2(2):267-272 [FREE Full text] [Medline: 18044143]
5. Piette JD, Rosland AM, Silveira M, Kabeto M, Langa KM. The case for involving adult children outside of the household in the self-management support of older adults with chronic illnesses. Chronic Illn 2010 Mar;6(1):34-45 [FREE Full text] [doi: 10.1177/1742395309347804] [Medline: 20308349]
6. Schoonover CB, Brody EM, Hoffman C, Kleban MH. Parent care and geographically distant children. Res Aging 1988 Dec;10(4):472-492. [Medline: 3227152]
7. Koerin BB, Harrigan MP. P.S. I love you. Journal of Gerontological Social Work 2003 Sep 24;40(1-2):63-81 [FREE Full text] [doi: 10.1300/J083v40n01_05]
8. National Alliance for Caregiving and the AARP. 1997. Family caregiving in the US: Findings from a national survey URL: http://assets.aarp.org/rgcenter/il/caregiving_97.pdf [accessed 2012-11-25] [WebCite Cache ID 6CTBmUfFH]
9. National Alliance for Caregiving and the Metlife Mature Market Institute. 2004. Miles away: The Metlife study of long-distance caregiving URL: http://www.caregiving.org/data/milesaway.pdf [accessed 2012-11-25] [WebCite Cache ID 6CTAg7AsI]
10. US Department of Health and Human Services, Agency for Healthcare Research and Quality. 2012. Health information technology: Best practices transforming quality, safety, and efficiency URL: http://healthit.ahrq.gov/portal/server.pt?open=514&objID=5554&mode=2&holderDisplayURL=http://wci-pubcontent/publish/communities/k_o/knowledge_library/key_topics/consumer_health_it/consumer_health_it_applications.html [accessed 2012-11-25] [WebCite Cache ID 6CTBxQqvr]
11. Baker LC, Bundorf MK, Singer S, Wagner TH. Validity of the survey of health and Internet and Knowledge Network's panel and sampling. 2003 May. URL: http://www.cdc.gov/pcd/issues/2004/oct/pdf/04_0004_01.pdf [accessed 2012-11-25] [WebCite Cache ID 6CTC6UJ3B]
12. Chang L, Krosnick JA. National surveys via RDD telephone interviewing versus the Internet: Comparing sample representativeness and response quality. Public Opinion Quarterly 2009 Dec 01;73(4):641-678. [doi: 10.1093/poq/nfp075]

13.   Knowledge Networks. 2010 Mar. KnowledgePanel: Processes & procedures contributing to sample representativeness & tests for self-selection bias URL: http://www.knowledgenetworks.com/ganp/docs/KnowledgePanelR-Statistical-Methods-Note.pdf

14.   Rosland AM, Heisler M, Janevic M, Connell C, Langa KM, Ker EA, et al. Current and potential support for chronic disease management in the United States: The perspective of family and friends of chronically ill adults. Families, Systems, and Health 2013:xx-xx (forthcoming).

15.   Zulman DM, Schafenacker A, Barr KL, Moore IT, Fisher J, McCurdy K, et al. Adapting an in-person patient-caregiver communication intervention to a tailored web-based format. Psychooncology 2012 Mar;21(3):336-341 [FREE Full text] [doi: 10.1002/pon.1900] [Medline: 21830255]

16.   DeSalvo KB, Fisher WP, Tran K, Bloser N, Merrill W, Peabody J. Assessing measurement properties of two single-item general health measures. Qual Life Res 2006 Mar;15(2):191-201. [doi: 10.1007/s11136-005-0887-2] [Medline: 16468076]

17.   Fox SA, Brenner J. Family caregivers online. Washington, DC: Pew Internet & American Life Project; 2012 Jul 12. URL: http://pewinternet.org/~/media//Files/Reports/2012/PIP_Family_Caregivers_Online.pdf [accessed 2012-11-26] [WebCite Cache ID 6CTEzEChs]

18.   National Alliance for Caregiving and United Healthcare. 2011. e-Connected family caregiver: Bringing caregiving into the 21st century URL: http://www.caregiving.org/data/FINAL_eConnected_Family_Caregiver_Study_Jan%202011.pdf [accessed 2012-11-26] [WebCite Cache ID 6CTLDtD51]

19.   Pagliari C, Detmer D, Singleton P. Potential of electronic personal health records. BMJ 2007 Aug 18;335(7615):330-333 [FREE Full text] [doi: 10.1136/bmj.39279.482963.AD] [Medline: 17703042]

20.   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]

21.   Zulman DM, Nazi KM, Turvey CL, Wagner TH, Woods SS, An LC. Patient interest in sharing personal health record information: A web-based survey. Ann Intern Med 2011 Dec 20;155(12):805-810. [doi: 10.1059/0003-4819-155-12-201112200-00002] [Medline: 22184687]

22.   California HealthCare Foundation. 2010. Consumers and health information technology: A national survey URL: http://www.chcf.org/~/media/MEDIA%20LIBRARY%20Files/PDF/C/PDF%20ConsumersHealthInfoTechnologyNationalSurvey.pdf [accessed 2012-11-26] [WebCite Cache ID 6CTLmRqFA]

23.   Walker J, Leveille SG, Ngo L, Vodicka E, Darer JD, Dhanireddy S, et al. Inviting patients to read their doctors' notes: Patients and doctors look ahead: Patient and physician surveys. Ann Intern Med 2011 Dec 20;155(12):811-819. [doi: 10.1059/0003-4819-155-12-201112200-00003] [Medline: 22184688]

24.   Cleveland Clinic. 2012. MyChart URL: http://my.clevelandclinic.org/online-services/mychart.aspx [accessed 2012-11-26] [WebCite Cache ID 6CTM5vNI9]

25.   Kaiser Permanente. 2012. My Health Manager: My Doctor URL: http://mydoctor.kaiserpermanente.org/ncal/Images/kporgbook_myhealthmanager_tcm28-74085.pdf [accessed 2012-11-26] [WebCite Cache ID 6CTMCQ5o0]

26.   GroupHealth. 2012. My GroupHealth URL: http://www.ghc.org/mygrouphealthpromos/onlinesvcs.jhtml [accessed 2012-11-26] [WebCite Cache ID 6CTMFmTPP]

27.   Feeley TW, Shine KI. Access to the medical record for patients and involved providers: Transparency through electronic tools. Ann Intern Med 2011 Dec 20;155(12):853-854. [doi: 10.1059/0003-4819-155-12-201112200-00010] [Medline: 22184694]

28.   Botek AM. Agingcare.com. 2012. 12 handy apps for caregivers URL: http://www.agingcare.com/Articles/12-Handy-Apps-for-Caregivers-149392.htm [accessed 2012-11-26] [WebCite Cache ID 6CTMZEU5y]

29.   The CareGiver Partnership. 2011. New tools for long-distance caregiving URL: http://blog.caregiverpartnership.com/2011/03/new-tools-for-long-distance-caregivers.html [accessed 2012-11-26] [WebCite Cache ID 6CTMdcGz2]

30.   Langa KM, Vijan S, Hayward RA, Chernew ME, Blaum CS, Kabeto MU, et al. Informal caregiving for diabetes and diabetic complications among elderly americans. J Gerontol B Psychol Sci Soc Sci 2002 May;57(3):S177-S186. [Medline: 11983744]

31.   Rosland AM, Piette JD, Choi H, Heisler M. Family and friend participation in primary care visits of patients with diabetes or heart failure: Patient and physician determinants and experiences. Med Care 2011 Jan;49(1):37-45. [doi: 10.1097/MLR.0b013e3181f37d28] [Medline: 21102357]

32.   Zickuhr K, Smith A. Digital differences. Washington, DC: Pew Internet & American Life Project; 2012 Apr 13. URL: http://pewinternet.org/~/media//Files/Reports/2012/PIP_Digital_differences_041312.pdf [accessed 2012-11-26] [WebCite Cache ID 6CTMpXCr3]

Abbreviations

CI: confidence interval
OR: odds ratio
PHR: personal health record

XSL•FO
RenderX

Original Paper

Major Infection Events Over 5 Years: How Is Media Coverage Influencing Online Information Needs of Health Care Professionals and the Public?

Patty Kostkova[1], Bc, MSc, PhD; David Fowler[2], BSc (Hons), MSc, PhD; Sue Wiseman[2], Registered Nurse (RN), MSc PH; Julius R Weinberg[3], MEd, MSc, DM, FRCP, FFPH

[1]Department of Computer Science, National Resource for Infection Control (NRIC), University College London, London, United Kingdom
[2]National Resource for Infection Control (NRIC), London, United Kingdom
[3]Vice Chancellor's Office, Kingston University London, London, United Kingdom

Corresponding Author:
Patty Kostkova, Bc, MSc, PhD
Department of Computer Science
National Resource for Infection Control (NRIC)
University College London
Gower Street
London, WC1E 6BT
United Kingdom
Phone: 44 20 7679 0340
Fax: 44 20 7679 0340
Email: P.Kostkova@ucl.ac.uk

Abstract

Background: The last decade witnessed turbulent events in public health. Emerging infections, increase of antimicrobial resistance, deliberately released threats and ongoing battles with common illnesses were amplified by the spread of disease through increased international travel. The Internet has dramatically changed the availability of information about outbreaks; however, little research has been done in comparing the online behavior of public and professionals around the same events and the effect of media coverage of outbreaks on information needs.

Objective: To investigate professional and public online information needs around major infection outbreaks and correlate these with media coverage. Questions include (1) How do health care professionals' online needs for public health and infection control information differ from those of the public?, (2) Does dramatic media coverage of outbreaks contribute to the information needs among the public?, and (3) How do incidents of diseases and major policy events relate to the information needs of professionals?

Methods: We used three longitudinal time-based datasets from mid-2006 until end of 2010: (1) a unique record of professional online behavior on UK infection portals: National electronic Library of Infection and National Resource of Infection Control (NeLI/NRIC), (2) equivalent public online information needs (Google Trends), and (3) relevant media coverage (LexisNexis). Analysis of NeLI/NRIC logs identified the highest interest around six major infectious diseases: *Clostridium difficile* (*C difficile*)/Methicillin-resistant *Staphylococcus aureus* (MRSA), tuberculosis, meningitis, norovirus, and influenza. After pre-processing, the datasets were analyzed and triangulated with each other.

Results: Public information needs were more static, following the actual disease occurrence less than those of professionals, whose needs increase with public health events (eg, MRSA/*C difficile*) and the release of major national policies or important documents. Media coverage of events resulted in major public interest (eg, the 2007/2008 UK outbreak of *C difficile*/MRSA). An exception was norovirus, showing a seasonal pattern for both public and professionals, which matched the periodic disease occurrence. Meningitis was a clear example of a disease with heightened media coverage tending to focus on individual and celebrity cases. Influenza was a major concern during the 2009 H1N1 outbreak creating massive public interest in line with the spring and autumn peaks in cases; although in autumn 2009, there was no corresponding increase in media coverage. Online resources play an increasing role in fulfilling professionals' and public information needs.

Conclusions: Significant factors related to a surge of professional interest around a disease were typically key publications and major policy changes. Public interests seem more static and correlate with media influence but to a lesser extent than expected.

The only exception was norovirus, exhibiting online public and professional interest correlating with seasonal occurrences of the disease. Public health agencies with responsibility for risk communication of public health events, in particular during outbreaks and emergencies, need to collaborate with media in order to ensure the coverage is high quality and evidence-based, while professionals' information needs remain mainly fulfilled by online open access to key resources.

Introduction

Background

There is a large amount of medical information available on the Internet, ranging from specialist databases and indexed collections of articles for health care professionals to less technical information sites for the general public. It is estimated that around 80% of the general public and a comparable proportion of medical professionals access information via the Internet [1]. In this paper, we examine the search behavior of visitors to a specialist medical online portal (in the domain of infectious diseases and infection prevention control) and the search behavior of the wider public using a search engine. We also consider the possible influence of media reporting of disease outbreaks on these behaviors.

The last decade witnessed turbulent events in the domain of infectious diseases and public health. New and emerging infections, such as Severe Acute Respiratory Syndrome (SARS), deliberately released threats (eg, anthrax), and ongoing battles with common illnesses, such as influenza, tuberculosis (TB), Healthcare Associated Infections (HAI), and the A/H1N1 swine flu pandemic outbreak of 2009 were amplified by the spread of disease through increased speed and volume of international travel. It is more important than ever to ensure that health care professionals and members of the public are well informed and kept up to date with the latest public health developments, government advice, and rapid risk communications. However, in addition to official health authorities' communications, in the Internet era professionals and the public increasingly use online resources to meet their information needs and seek up-to-date evidence. Also, media coverage of infection outbreaks, public health issues, and media-mediated risk advice is increasingly influencing public perceptions and often distorting health critical information [2].

Health-Related Information Seeking Behavior of Professionals and Public

As of the end of December 2009, there were an estimated 1.8 billion Internet users worldwide. In Europe, 53% of the population use the Internet, which rises to 77% in the United Kingdom (with 69% having a broadband connection) (values are from surveys quoted in Higgins et al [1]). More recent results [3] indicate that 56% of the population in the European Union use the Internet daily, with 68% using the Internet every week.

Various surveys ([4-5]) quoted by Higgins et al [1] indicate that 8 out of 10 Internet users in the United States use the Internet to access health information and that the corresponding number for Europe was 7 out of 10 (according to a 2007 study by Andreassen et al [6]). A study by Seybert in 2011 [3] found that 54% of EU Internet users used the Internet to look for health-related information (lower than the 71% mentioned by Andreassen et al). This difference might be explained by differences in sampling and the wording of questions (see [7] for a discussion on this subject). Overall, it seems reasonable to expect a continued increase in the proportion of Internet usage by the general public, as well as the proportion of those users seeking online health information.

In addition to the increased use of online resources by members of the public to manage their personal health and better understand their conditions, in recent years the online health information environment has become mobile, with 17% of cell phone users having used their phones to look up health information and 9% using software applications on their phones that help them track or manage their health [8].

While these studies provide cumulative data on Internet usage, it is also essential to investigate users' search and online behavior to understand their online information needs and how these are fulfilled technically as well as in the context of site usability [9]. Furthermore, do members of the public access medical information online for the same reasons as health care professionals and does their search behavior differ?

A number of studies have investigated health care professionals' online information seeking behavior. Younger gives a survey of studies comparing the search behavior of doctors and nurses [10]. It was difficult to compare individual studies due to the lack of harmonization of design and terminology, but the main conclusion was that many barriers exist for health care professionals, including lack of time and resources. There are also social barriers for professionals to use computers in the health care environment [11]. Alghanim [12] examines the information seeking behavior of primary health care physicians in Saudi Arabia, with one finding showing that around 50% of rural physicians used online databases and general websites to find information, rising to over 70% for urban physicians, with the difference presumably due to the lack of availability of these resources in rural areas. O'Keeffe et al [13] surveyed the information seeking behavior of a variety of health care personnel at two medical establishments in northern California (however the study does not distinguish exactly between online and offline information).

Public health and infection is one of the most varied domains of medicine, subject to rapid changes, disease outbreaks, and control measures involving the general public at regional, national, and international scales. As we run a specialist online digital library for infection and public health professionals and

have a unique longitudinal online search dataset, we will focus on the information needs of the public and professionals regarding infection. It is not easy to say what drives the behavior of the public to seek information on particular infectious diseases. An actual outbreak of the disease could be a factor, but the knowledge of the outbreak will usually be obtained via mass media. The media's reporting of disease outbreaks may be exaggerated due to certain needs, such as a need for a human interest angle [14].

Effects of Media Coverage on Information Seeking Behavior

Media coverage of health-related news stories can influence the decisions and behavior of policy makers and the public [2]. For example, some parents refused to have their children vaccinated with the combined MMR (measles, mumps, and rubella) vaccine after intense media coverage of a single paper (later discredited [15]) linking the MMR vaccine to autism. Media coverage can be distorted, giving extra attention to stories about health concerns that have little real impact, while largely ignoring those (such as smoking, obesity, and alcohol) that cause much more harm [2].

Media coverage can also help to limit an outbreak, by causing individuals who are susceptible to the disease to isolate themselves from infected individuals [16]. Finally, the media coverage of a disease may be heightened even when there is no outbreak at the time. A good example is the reporting of the findings of an inquiry into an outbreak (eg, the coverage in October 2007 of *Clostridium difficile* (*C diff*), concerning a report into a prolonged outbreak between April 2004 and September 2006).

The 2009 swine flu outbreak was a health event covered extensively by the mass media and with an impact investigated by a number of research studies. Hilton and Hunt examined UK newspaper coverage of the 2009-10 swine flu (A/H1N1) outbreak [17]. They found that there was "immense" coverage in the spring and summer of 2009, when there was most uncertainty about the future impact on the United Kingdom. Later, in the autumn of 2009, there were few news articles, despite a second peak in the number of swine flu cases. Also, public information needs changed as members of public were overwhelmed by the information in the spring of 2009 but were less interested in the second half of 2009 [18].

Therefore, in this study we will investigate the following questions:

1. How do health care professionals' online search needs around infection differ from the needs of the public?
2. Does media coverage contribute to the information needs among the public for infection?
3. How are incidents of a disease and major policy events related to the information needs of professionals?

Methods

We used 3 time-based datasets that were selected to cover the levels of interest in various infectious diseases and organisms. The datasets are intended to give a good representation of the search interests of health care professionals and the public and also the level of media coverage of each topic. We were not attempting in this study to prove causal relationships between the highly interrelated worlds of public, professionals, and media coverage, but rather to use a triangulation method [19] to examine the 3 related datasets and seek to make inferences about possible causal relationships between them.

Datasets

The 3 time-based datasets are:

1. The levels of user activity for various infection topics in the NeLI/NRIC specialist online digital library, run by City eHealth Research Centre (CeRC), City University, London. This unique dataset reflects the levels of interest in various topics by health care professionals.
2. The search statistics for the same infection topics from Google Trends [20]. This dataset reflects the levels of interest in the topics by the general public who seek health care information online on Google.
3. The numbers of news articles retrieved from the LexisNexis database, concerning the same topics as were used for the other datasets. This dataset represents the media coverage of the topics. Our search was restricted to English language coverage, but this includes major world newspapers in English.

We were interested in trends in the levels of activity in these datasets (whether activity was above or below the average level, and by how much, and whether activity was rising or falling over the long term or showing sudden peaks) and any correlations between the datasets.

The datasets are described in more detail in the following sections. As our primary interest was professional needs and their correlations, the most reliable results were ensured by selecting the diseases and conditions that had the highest activity levels on the NeLI/NRIC sites.

Table 1 gives the average weekly and peak NeLI/NRIC category accesses for various diseases or organisms, arranged in descending order. Unsurprisingly, as the user base of NeLI/NRIC is predominantly infection control professionals and the government nationally has increasingly focused on targets to reduce the top two infections listed below (Table 1), *C difficile* and multi-resistant *Staphylococcus aureus* (MRSA) lead the table followed by tuberculosis, meningitis, norovirus, and influenza.

Table 1. The average and peak weekly accesses for various diseases/organisms in NeLI/NRIC.

Disease/organism	Average weekly accesses	Peak weekly accesses
C difficile + MRSA	43.3	N/A
C difficile	23.9	72
MRSA	19.4	54
Tuberculosis	13.9	51
Meningitis	13.2	133
Norovirus	6.3	74
Influenza	2.6	21
SARS	1.4	34

In our analysis, we decided to combine the results for *C difficile* and MRSA. This was because (1) they are related topics (health care associated infections), (2) they are often mentioned together in media articles, and (3) public searches for "*Clostridium difficile*" were very few in comparison to searches for "MRSA". This could possibly be due to the difficulty of spelling "*Clostridium difficile*" compared to "MRSA". For this reason, we also looked at public searches for "superbug" because this lay term was frequently used in media and covers all HAIs. Another possibility is that the UK government targeted MRSA reduction first and only much later targeted *C difficile.*

The timeframe of the study was from week 31 (end of July) 2006 until the end of 2010, which is the period for which we have NeLI/NRIC data. The other datasets (Google Trends data and LexisNexis news article data) also cover this period.

Dataset 1: Professional Information User Needs—the NeLI/NRIC Portal Dataset

It is hard to determine the information needs of health care professionals. While surveys of behavior have been performed, the NeLI/NRIC server logs contain an invaluable record of actual search behavior in the domain of infection over several years.

Initially the specialist Library of Infection [21], part of the National electronic Library of Health (later NHS Evidence), The National electronic Library of Infection (NeLI) [22] is an online digital library created in 2000 at CeRC, with the aim of bringing together the best available evidence-based resources on the investigation, treatment, prevention, and control of infectious disease [23] (see Figure 1). Under the NeLI umbrella, several projects were developed using the same model, the largest of which is the National Resource for Infection Control (NRIC)[24], which was set up in May 2005 and specializes in resources on infection control and prevention. In the rest of the paper, we consider NeLI and NRIC together.

In addition to providing up to date evidence-based resources and stating the level of evidence of each resource (RCT,

Meta-analysis, etc), a key benefit of NeLI/NRIC is that *Reviewer's Assessments* (RAs) are attached to documents within the library. These are written by professionals in the field and provide a short summary of what the document is about, highlighting any contradictory studies, potential bias, or conflicts of interest. Each review is signed by the reviewer and may be commented on by registered users. Documents can be found by searches (either simple keyword search or more complicated searches with Boolean operators) or by using a navigation structure based on a taxonomy of the domain developed with domain experts. Documents are organized in a two-level taxonomy that is also used for document indexing by domain experts and then further subdivided until documents about specific presentations, organisms, and diseases are found. In NeLI, the highest level categories are (1) Clinical Presentation, (2) Organisms, (3) Diseases, and (4) Systems, while in NRIC, the top level consists of (1) Settings, (2) Clinical Practice, (3) Transmission, (4) Diseases/Organisms, and (5) Policy/Guidance.

The source of professional health care interest data for our study is the Web traffic logs for NeLI/NRIC that have been automatically recorded since 2005. (Logs were recorded between 2001 and 2005 for NeLI, but as the site architecture changed in 2005, detailed comparisons between the logs for time periods before and after the change are not really possible).The Web server keeps a log of all Web accesses, and this record has been preserved since 2005.

Each log entry contains details of an HTTP request sent from a Web browser to the Web server, including the date of each request, the IP address of the visitor, the page requested, and other data. Figure 2 shows a typical entry, and Figure 3 shows the same entry with explanations of the fields.

We can only observe visitors' interactions with the NeLI/NRIC sites, with the notable exception that we can often determine which previous page they browsed before arriving at NeLI/NRIC (the *referring page*) that provides valuable information about user navigation behavior and successful promotion of the online resource.

Figure 1. The National electronic Library of Infection (NeLI; information can be found by using drop down menus, left, or the search box, right).
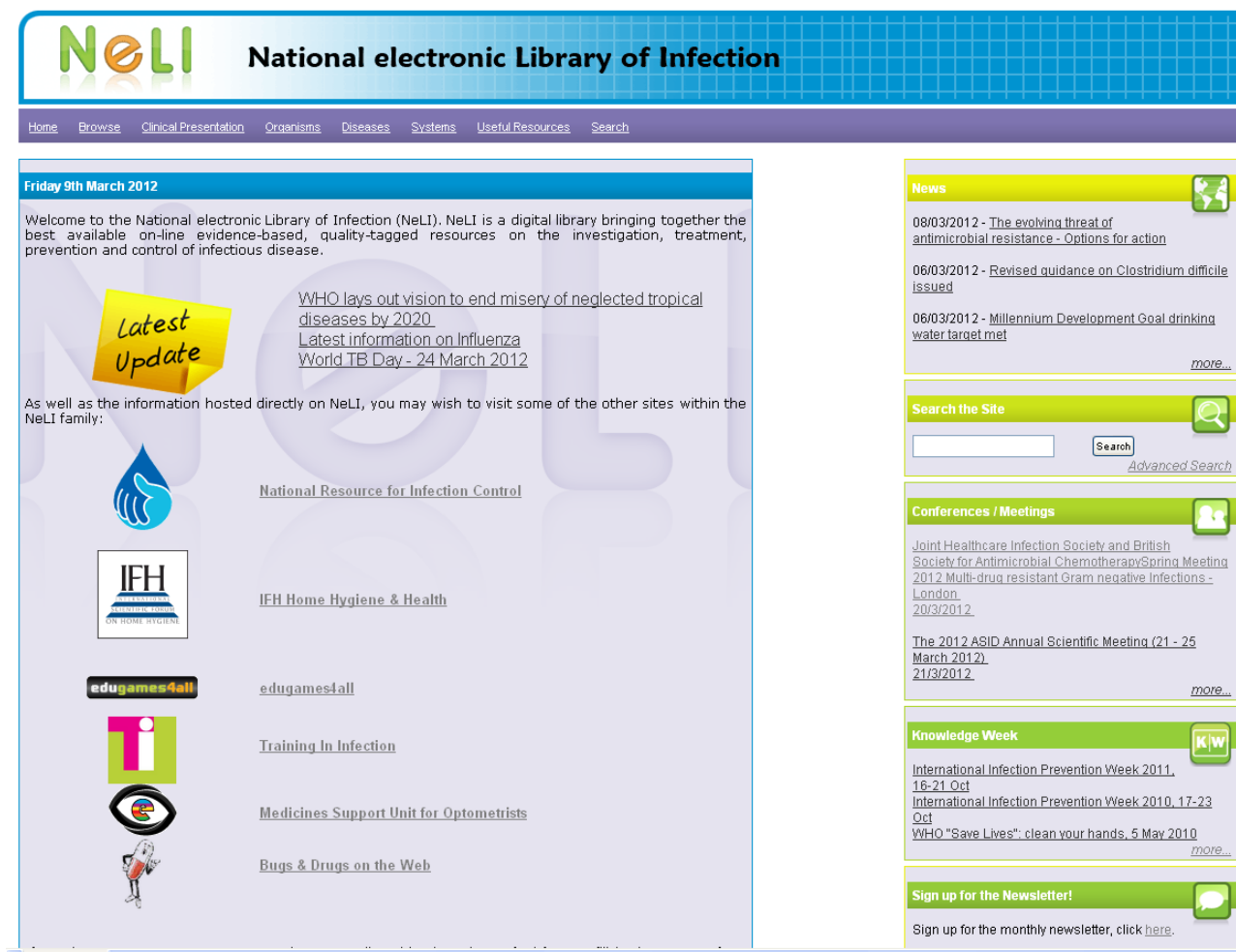
Figure 2. A sample log entry for a Web access to NeLI.

Figure 3. A sample log entry for a Web access to NeLI, annotated with the meanings of the available fields.

NeLI/NRIC Users

The NeLI/NRIC portals are aimed at health care professionals with interests in infection. Initially part of the NHS-led project, they were also promoted through the Health Protection Agency (HPA), the national public health agency in the United Kingdom. The site had over 5000 unique users per month in 2011 and between 20,000 and 30,000 page views (Figure 4).

From 2006 to 2008, NeLI/NRIC was heavily promoted at conferences and at other events, seemingly leading to an increase in visitor numbers. More recently, due to lack of resources, the site has been kept up to date, but promotional activity has lessened, resulting in a decline in site activity, clearly visible in the graph.

NeLI/NRIC is visited most frequently by users in the United Kingdom and the United States, and English speaking countries. However, despite the content being in English only and forming part of a national library, there is a growing number of users from countries such as India, Germany, and China, indicating

a global need for such an evidence-based open access portal (see Supplementary Figure 1 in Multimedia Appendix 1).

All the content of the evidence-based library is in the public domain and free to use. In order to improve accessibility and usability no registration is required to access the content although users can subscribe to receive a monthly electronic newsletter that highlights the latest resources and upcoming events and conferences. Users can join the subscription list either personally at a conference, at a study day where NeLI/NRIC is presented, or online at a dedicated subscription page. The subscription database holds details for over 3500 NeLI/NRIC users. Subscribers listed in the database can provide their professions and specialities. Although the primary interest in subscribing to the site is "infection", in order to better understand the professional backgrounds of NeLI/NRIC users an analysis of these was performed, and the breakdowns of professions and specialities are detailed in Supplementary Figures 2 and 3 in Multimedia Appendix 1.

Figure 4. The numbers of visitors and page views for NeLI/NRIC between 2006 and 2011.
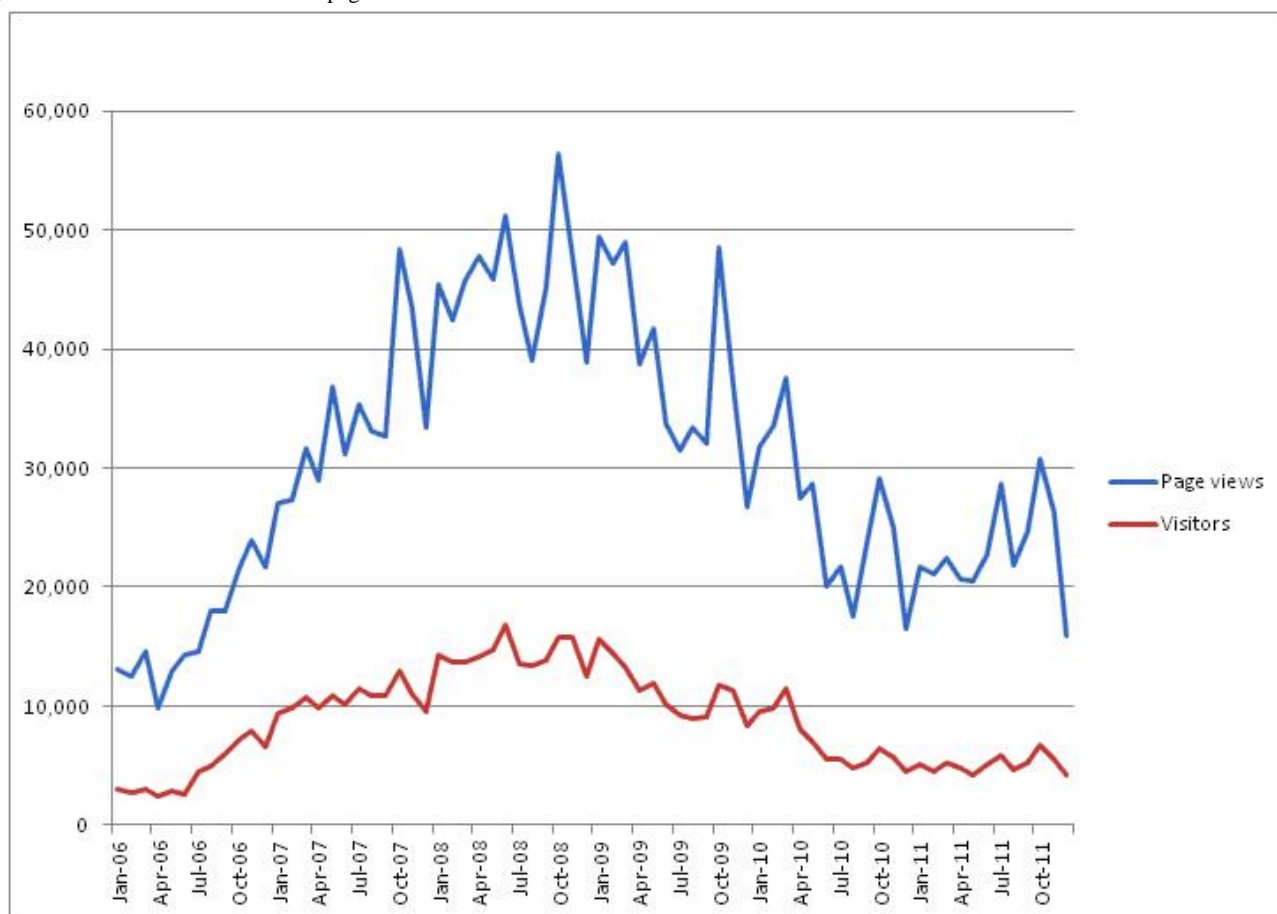
Dataset 2: Public Information Needs—the Google Trends Dataset

Although there are many public-facing websites about infection (eg, NHS Choices [25], Bugs and Drugs on the Web [26]), these are multiple and fragmented, some focused on a single condition (eg, Bugs and Drugs) and of varying quality as these are run by patient groups, governments, and industry [27,28]. Furthermore, the search logs are not publicly available. For this reason, using

search engine data for searches for infection-related terms provides a high volume and much more compelling source of public online information needs in this area over the same period of time. Therefore, to evaluate patient information needs, we used data from Google Trends [20] (Supplementary Figure 4 in Multimedia Appendix 1), which measures Google searches for particular keywords. These data measure the weekly volume of searches using a keyword, but rather than the absolute numbers of searches, a normalized value is given. This is scaled

so that for a single keyword or phrase, the average value over the specified time period is 1. Therefore, a value of 2 would indicate a volume of twice the long-term average. Comparisons between terms can also be made. In this case, the normalization is done so that one term has a long-term average of 1, and the other terms have values that are scaled accordingly.

Google Users

Google is currently the most commonly used search engine worldwide, with 90% of the market share globally, and 80% in the United States, according to StatCounter Global Stats [29]. See Supplementary Figures 5 and 6 in Multimedia Appendix 1 for the worldwide and US data respectively. It therefore seems justifiable to use Google search data as representative of the general public's search interests.

Dataset 3: Media Coverage of Infection Outbreaks—News Articles From LexisNexis

The third dataset, measuring media coverage of specific topics, is the newspaper articles retrieved from the LexisNexis database [30] (see Supplementary Figure 7 in Multimedia Appendix 1). These articles were from major world newspapers in the English language. The results can be saved as a text list, from which the dates of the articles (which are necessary for our analysis) can be extracted.

Analysis: Pre-Processing of NeLI/NRIC Log Data

Before Web server log data could be used, the dataset had to be cleaned. As there are site visits not motivated by specific interest that can be of high volume and at random times, it is important to try to identify and remove them from our data. The main sources of spurious accesses were *Web crawlers* and *referrer spam* (there are also accesses by the website developers during developing and testing, which were easily identifiable).

Web crawlers (also referred to as spiders) are programs that visit pages of a website, usually for the purpose of indexing the site for search engines. Web crawlers tend to visit the same sites frequently to check for updates. Crawlers can cause serious distortion of the Web log statistics, as they can produce a spike in the logs that is not due to any genuine interest in the site [31,32].

Referrer spam [33] is created by automated programs that generate Web log entries with the referrer site field set to a specific Web address. This is intended to generate free advertising if the weblogs are made available online.

In order to remove as many spurious log entries as possible, the Web logs were pre-processed with the following steps:

1. All entries with an IP address in a list of developer IPs were removed.
2. The browser type field (see Figure 4 for an example) in each log entry was examined and those that did not correspond to common Web browsers were removed. Our aim was to remove those entries that were not caused by human use of a Web browser.
3. The previous step still left a large number of entries that were clearly produced by Web crawlers. The remainder of the browser type field was examined, and any that contained certain keywords that indicate Web crawlers (specifically "bot", "crawler", "spider", "slurp", and "jeeves") were rejected.
4. Referrer spam was removed by first finding the most common referrer websites in our logs. By concentrating on the most frequently occurring sites, the most likely referrer spam sites were identified manually. A block list of terms and site names was built up, which was used to exclude log entries during the processing phase.

Analyzing Interest in Infection Topics

After the pre-processing that was only required for the professional needs containing NeLI/NRIC logs, the analysis of the 3 datasets was performed—each required a different technique to analyze an information need or interest in a certain infection topic.

NeLI/NRIC Logs

After pre-processing the logs, the remaining entries were divided into document views (ie, looking at a specific document in the library), category browses (looking at a list of documents about a specific topic, the second level of the two level taxonomy described in section 3.1.1), and searches (ie, the entry of search terms into the search box). Other accesses, such as image views or views of pages not relating to specific diseases or organisms, were not counted. We concentrated on category browsing and document views, as (1) browsing was much more commonly performed than searching (93% of the total) [34], (2) factors such as misspellings, synonyms, and complex search phrases make analysis of search terms much more complicated, and finally (3) for advanced keyword searches, an autocomplete function (which suggests keywords after the user has typed a few characters) was used. This distorts the results for searches, as the searches for partial words are also recorded in the logs.

Finding information on NeLI/NRIC can be achieved either by using a search (either from an external search engine or by an internal search on the NeLI/NRIC site) or by visiting the site and following links or menus to navigate to the required information (see Figure 5).

After plotting graphs for professional interest in infectious diseases using the NeLI/NRIC dataset, we considered the major peaks in the graphs and attempted to identify any major policy documents that were released at the corresponding times. To verify if these documents may have caused the peak in interest, we studied the NeLI/NRIC logs to measure the download rates of the documents and matched them to the professional interest graphs (which measure the overall browsing activity for the disease).

Figure 5. Alternative methods of finding information on NeLI using the drop down menus (browsing) and free text searching.
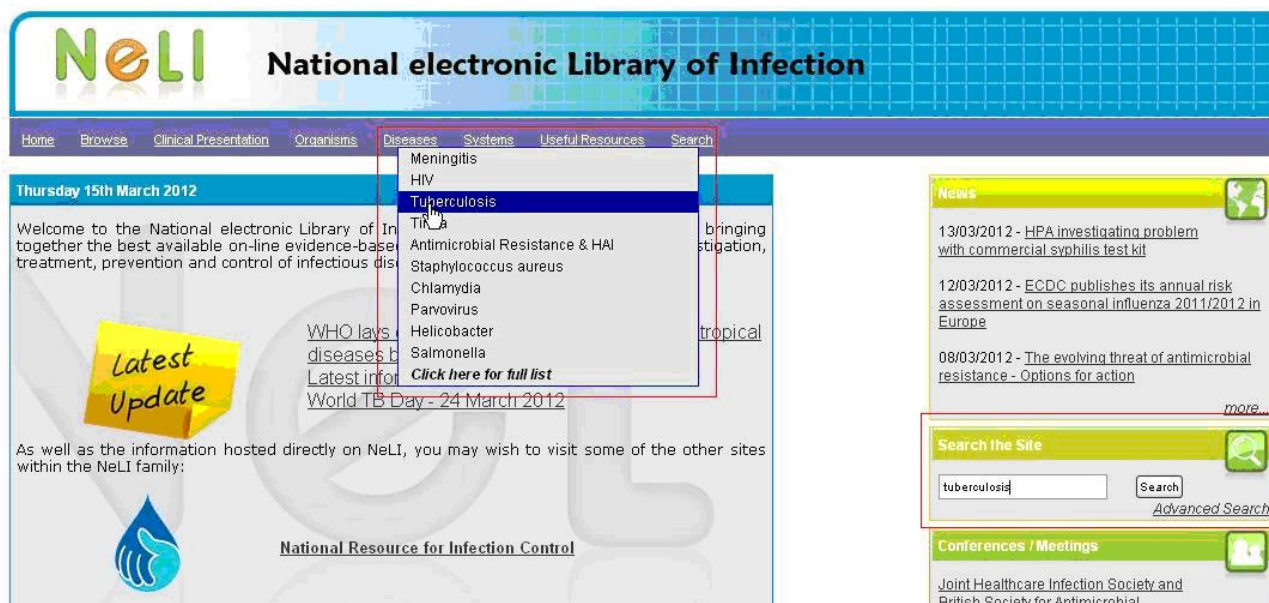
Public Interest: Google Trends Dataset Analysis

Public interest data were obtained by entering search terms at the Google Trends website. The data were downloaded as weekly data in comma separated value (CSV) format, using relative scaling, so that the data are scaled to make the average level over the period is 1. We chose to use single terms (eg, "norovirus", "tuberculosis", "*C difficile*") instead of trying to include synonyms ("winter vomiting bug", "TB", "C. diff", "superbug", etc), partly for simplicity and partly because Google's search algorithm can already make some allowance for synonyms and misspelled search terms. There are specific complications associated with analyzing influenza, as in addition to the common term "flu", there are varieties of influenza that have been widely covered by the media (avian influenza, or "bird flu", and swine influenza, or "swine flu"). As it is difficult to tell whether the results for "influenza" might have been part of a more specific query about avian or swine influenza, we found totals for "influenza" as a whole.

Media Coverage: Newspaper Article Analysis

Media articles were extracted from the LexisNexis database, using the same search terms as were used for the Google Trends results. The search was performed over "Major World Publications (English)" and returned articles where the keyword was mentioned near the start of the article. The similarity measure was set to "On, high similarity" to exclude duplicate articles. The articles were sorted by date and counted to give weekly totals.

Once the information from the 3 datasets was plotted as time series, we examined the correlation of the 3 signals and further investigated the real events and key publications to attempt to explain any spikes, trends, and other patterns in the data. These were gathered by searching news and press release databases from the Health Protection Agency (HPA), the European Centre for Disease Prevention and Control (ECDC), the World Health Organization (WHO) and other agencies, and in-depth analysis of access to the actual documents on NeLI/NRIC creating the

peak and manually verified by infection domain experts. Also, we related the levels of professional interest with that of the public and each of these to the media coverage.

Results

To compare the 3 datasets, we applied the same scaling that Google Trends uses to the numbers of news stories and levels of professional interest, so that the average level (or "baseline") over the study period was 1. This means that the 3 measures could be plotted on the same scale.

Graphs were plotted showing the activity over time for NeLI/NRIC and Google Trends. Google Trends normalizes its data so that a level of 1 is the long-term average activity level over the period, and for comparison the same was done with NeLI/NRIC.

Major Infection Outbreaks

The last decade has been eventful in the domain of infectious disease. There have been periods of emerging infections (SARS), epidemic outbreaks (avian flu), a pandemic (swine flu in 2009) as well as recurring outbreaks for common infections (influenza, MRSA). We evaluated the 3 datasets to try to understand the correlations; however, as the primary aim was to understand professionals' needs, we investigated the NELI/NRIC dataset to determine the most accessed infection topics and disease outbreaks.

The next section describes the results of each disease separately and provides background events to illustrate the information needs of professionals and public.

Clostridium difficile **and MRSA**

Introduction

Clostridium difficile, also written as *C difficile* or *C diff*, and MRSA are bacteria that can infect patients through cross infection, in hospitals, nursing homes, or other health care facilities, hence the commonly used term health care associated

infections (HAIs). They are also linked to overuse of antibiotics causing resistance or damage to normal body bacteria, poor hygiene practice, age, and lowered immunity. Another popular term among the public is "superbugs" [35,36].

Table 2 shows the number of news articles returned from searches on "*Clostridium difficile*" and "MRSA" and the articles that appear in both sets of results. There are relatively few articles (11.3%) that are about *C difficile* alone, whereas there are many more (62.3%) on MRSA alone, with about a quarter of articles (26.4%) mentioning both HAIs. Figure 6 shows a comparison of Google searches for "*Clostridium difficile*" and "MRSA" (and the term "superbug" for reference). Again, MRSA is a more popular term, with around ten times the number of searches performed. This might be partially explained by the difficulty of spelling "*Clostridium difficile*" compared to "MRSA" when searching. There is also the possibility that the public were more alarmed about MRSA, and there was a strong public support network (including MRSA action groups) bringing it to the public's attention.

Results

Figure 7 shows the levels of (1) professional interest (measured by numbers of NeLI/NRIC accesses of the *Clostridium difficile* and MRSA taxonomy pages), (2) public interest (measured by comparative frequencies of Google searches for the terms "*Clostridium difficile*" and "MRSA"), and (3) media coverage (measured by the number of news articles mentioning "*Clostridium difficile*" or "MRSA" obtained from the LexisNexis database). Each statistic is measured weekly and normalized so that the baseline average over the period is 1.

The first observation (which applies generally to other diseases/organisms) is that the professional interest (measured by NeLI/NRIC accesses) is "noisier" (have a higher variance) than the Google Trends data. This is clearly due to Google's far larger traffic volume. It is difficult to get exact figures for Google's search volume, but using Google's AdWords service indicates that in the year to January 2012, the average monthly global number of searches for the phrase "what is *C diff*" was 368,000 [37]. If other searches related to *C difficile* were to be included, the total number of relevant searches would be much higher.

Professional Interest

The maximum level of professional interest occurs at week 43 in 2007 (which also coincides with the maximum levels of public interest and media coverage). The professional interest level at this peak was 2.6 times the baseline level, compared to 5.0 times the baseline for media coverage, and 5.7 times for public interest. There are periods of increased activity early in the second half of 2006 and in the first half of 2010. It is likely

that the increased activity in 2006 is due to the high levels of promotional activities for the newly relaunched NeLI/NRIC sites. More evidence for this comes from comparing the graphs in later sections, which show a similar pattern. The heightened activity in 2010 may be due to promotional activities or to the aftermath of pandemic flu, but this is unclear.

Public Interest

The public interest shows a very clear spike in 2007, coinciding with the spikes in the professional interest and media coverage. This spike is at a level that is 5.7 times the baseline. There is also a slight dip in the interest level at the end of each year, which can also be seen in the graphs for most of the other diseases. This is presumably due to lower levels of search over the period of the Christmas and New Year holidays. Also, overall public interest decreases after the 2007 spike until the end of the study period, where it is at a similar level to before 2007.

A possible interpretation of the public interest in MRSA and *C difficile* is that in 2007 it was affecting them, their relatives, and friends. In addition (in the United Kingdom at least) MRSA action groups were very active at this time and public pressure finally made the government take action, introducing targets for MRSA reduction in hospitals and nursing homes. Subsequently MRSA incidence fell and *C difficile* incidence increased, before attention turned to tackling *C difficile*.

Media Coverage

The media coverage also shows a clear peak in late 2007. The news stories at this time mainly focused on the findings of the Healthcare Commission in the United Kingdom concerning an outbreak of *C difficile* between April 2004 and September 2006. For example:

> *Scores of NHS patients were killed during Britain's deadliest outbreak of a hospital superbug, a damning report by the government's health watchdog reveals today. The Healthcare Commission attributed the deaths of 90 patients at the Maidstone and Tunbridge Wells hospitals in Kent to infection from C difficile, which causes severe diarrhoea and has taken over from MRSA as the main threat to patients.* [38]

This indicates that heightened media coverage in late 2007 over the Healthcare Commission report had a correlation to professionals' needs, who were likely to access the report but around 3 times rather than 6 times more frequently than the baseline.

The article is also an example of MRSA being mentioned in a report that is mainly about *C difficile*.

Table 2. Number of news stories from 2006-2011 about *C difficile* and MRSA and the number of articles common to both lists.

Keyword	"*Clostridium difficile*" only	"MRSA" only	Both terms
Number of articles mentioning the term	197	464	138
Percentages	11.3	62.3	26.4

Figure 6. A comparison of Google searches for the terms "clostridium difficile", "MRSA", and "superbug".
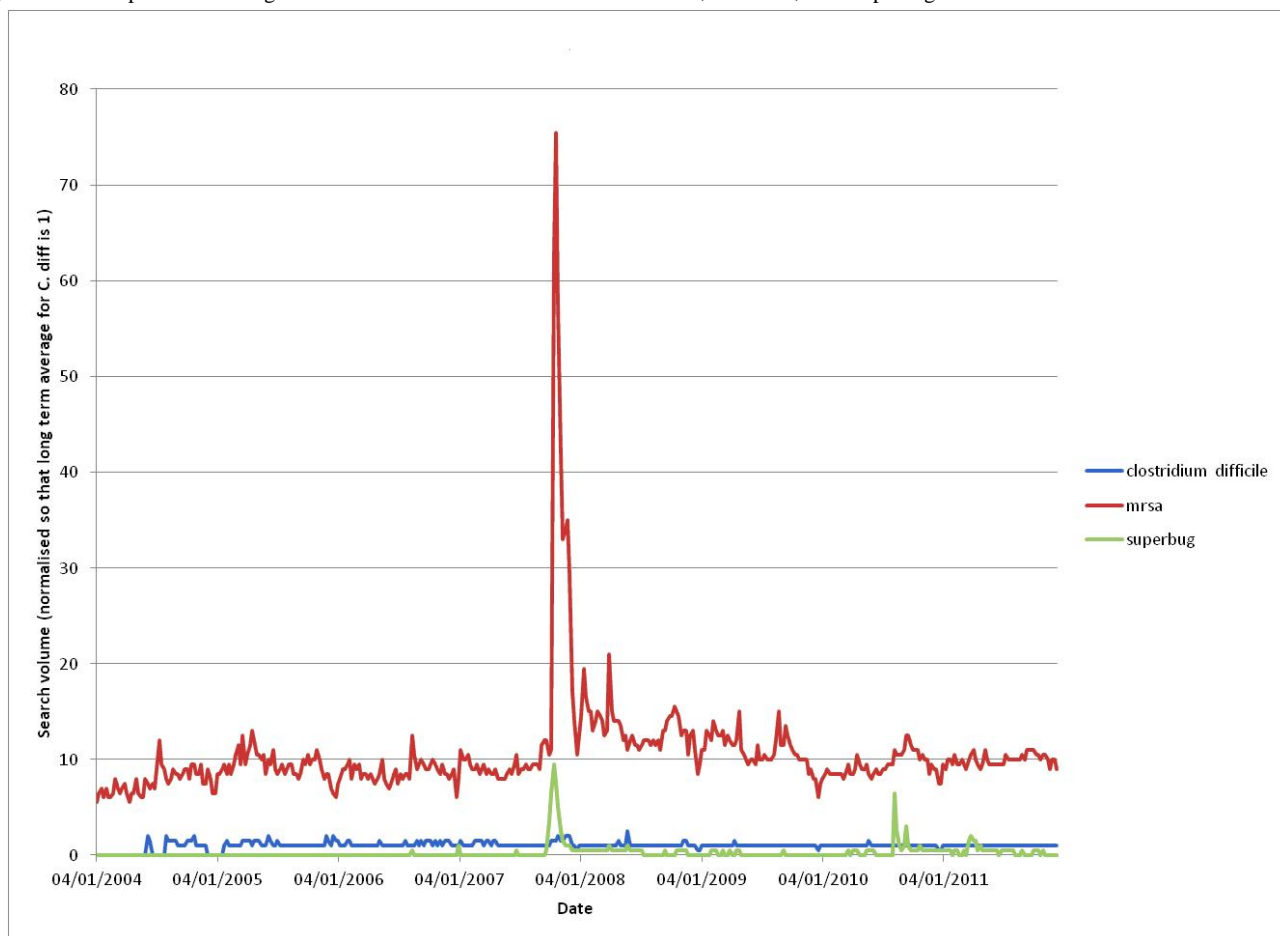
Figure 7. The public and professional interest, and media coverage for *Clostridium difficile* and MRSA.
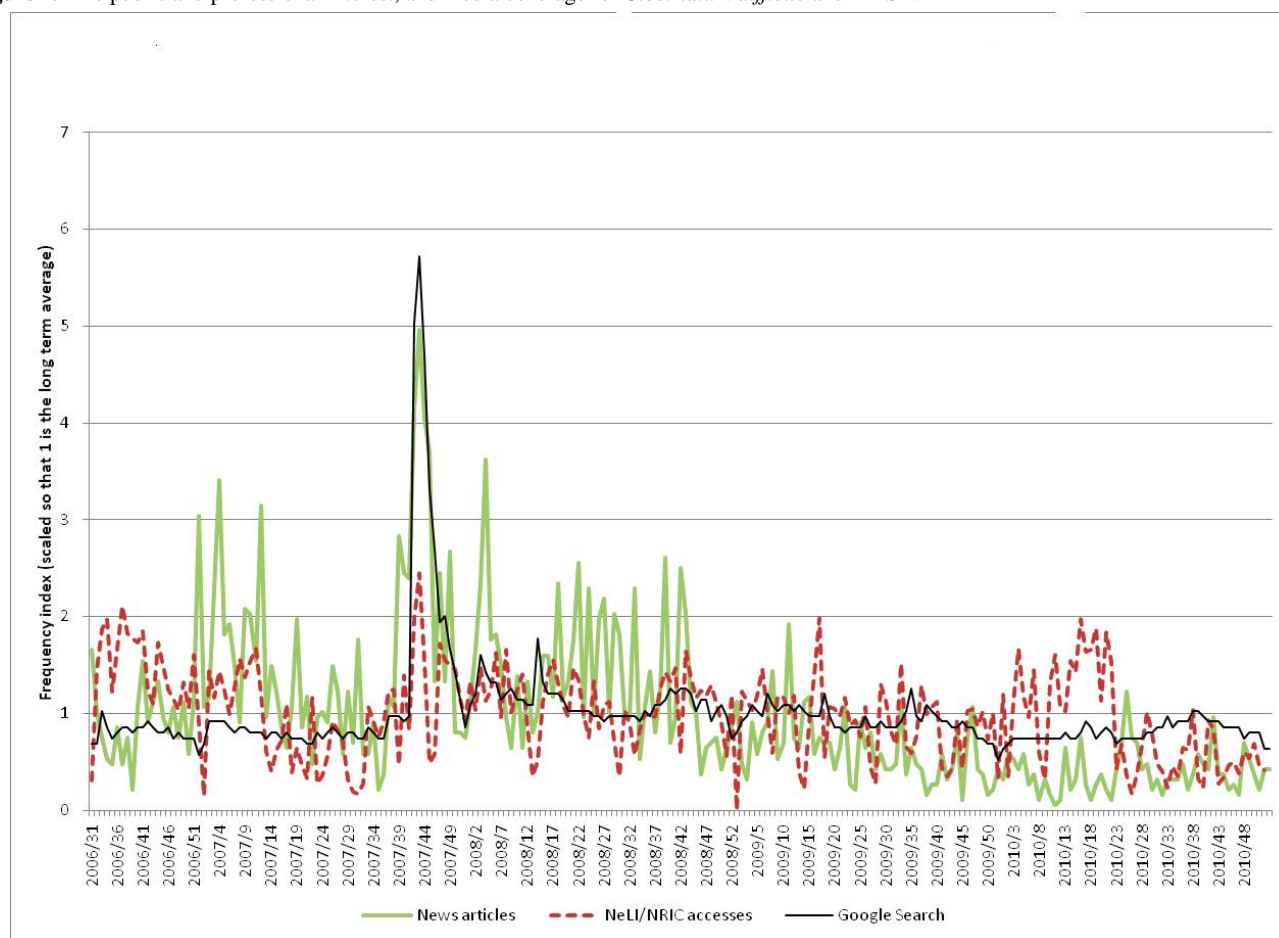
Tuberculosis

Introduction

Tuberculosis (TB) is an infectious disease that is caused by a bacterium called *Mycobacterium tuberculosis*. TB primarily affects the lungs, but it can also affect organs in the central nervous system, lymphatic system, and circulatory system. Infection is spread when bacteria, coughed up by an individual with TB affecting their lungs, are released into the air and inhaled by others. TB is a major global health problem and also prevalent among people with HIV/AIDS.

Results

From the graphs for TB (see Figure 8), it seems clear that professional and (to a lesser extent) public interest are both declining gradually. It is therefore not surprising that the NeLI/NRIC levels of interest did not show corresponding peaks or that Google Trends does not have a peak (as tuberculosis is not central to the stories, users would not use the term in searching for content).

Professional Interest

The overall level of professional interest appears from the graph to be declining, although this may be misleading. The graph shows the same higher level of interest for the first few months of the study period, which coincides with the promotional activities that would generate higher activity levels from

professionals. When this is discounted, the remaining interest levels are more level.

A tuberculosis "Knowledge Week" was held on NeLI/NRIC in conjunction with the HPA from March 26-30, 2007, to provide health care professionals with quick and easily accessible up-to-date knowledge on the disease. This activity does not appear as a peak in the graph, as the most accessed page for the Knowledge Week was a special front page that was not counted in our analysis of searches for the disease.

A document "Tuberculosis prevention and treatment: a toolkit for planning, commissioning and delivering high-quality services in England" was published by the NHS on June 15, 2007, as TB was becoming a growing and expensive problem in the United Kingdom. This shows up as a peak (3.2 times baseline) in week 26 of 2007, following a public interest peak in week 21 (2.8 times baseline), a professional interest peak (4.2 times) in week 22, and a media peak (5.7 times) in week 23. Figure 9 shows the professional interest (the same NeLI/NRIC accesses as in Figure 8) for tuberculosis, together with the accesses for the document (measured as a proportion of the total weekly document accesses). Clearly there is a surge of interest in the document at the time of publication (week 26), followed by a steady decline.

Public Interest

The highest level in public interest occurred in May 2007 (week 21, 2.8 times baseline), which seems to be related to a story

concerning a US citizen infected with TB who took a flight to Europe, potentially spreading the disease [39,40]. On this occasion, public interest was most likely triggered by the media story. The yearly dip in interest at the end of each year is also evident.

Tuberculosis is not highly prevalent in the United Kingdom but has been increasing and affects immigrants and the homeless more than other groups. This may explain some of the lack of public interest, as these affected groups may not have as much access to the Internet as other members of the public.

Media Coverage

There are several peaks in the media coverage, notably in mid-2007 (5.7 times baseline, mainly relating to a long-running

story about a bullock kept at a Hindu temple that contracted TB [41]), mid-2008 (1.9 times baseline, mainly about a potential cull of badgers to control bovine TB [42]), and late 2008 (about a successful human windpipe transplant, which was needed due to the patient's earlier case of TB [43]). These media stories tended to be about animal (specifically bovine) TB and a single human interest story, where the disease was incidental. Finally, there are peaks each year from 2007-2010 (2.5, 2.7, 3.1, and 3.1 times the baseline, respectively) coinciding with the World TB day, which is March 24 each year. This indicates that such events can generate media coverage.

Figure 8. The public and professional interest, and media coverage for Tuberculosis.
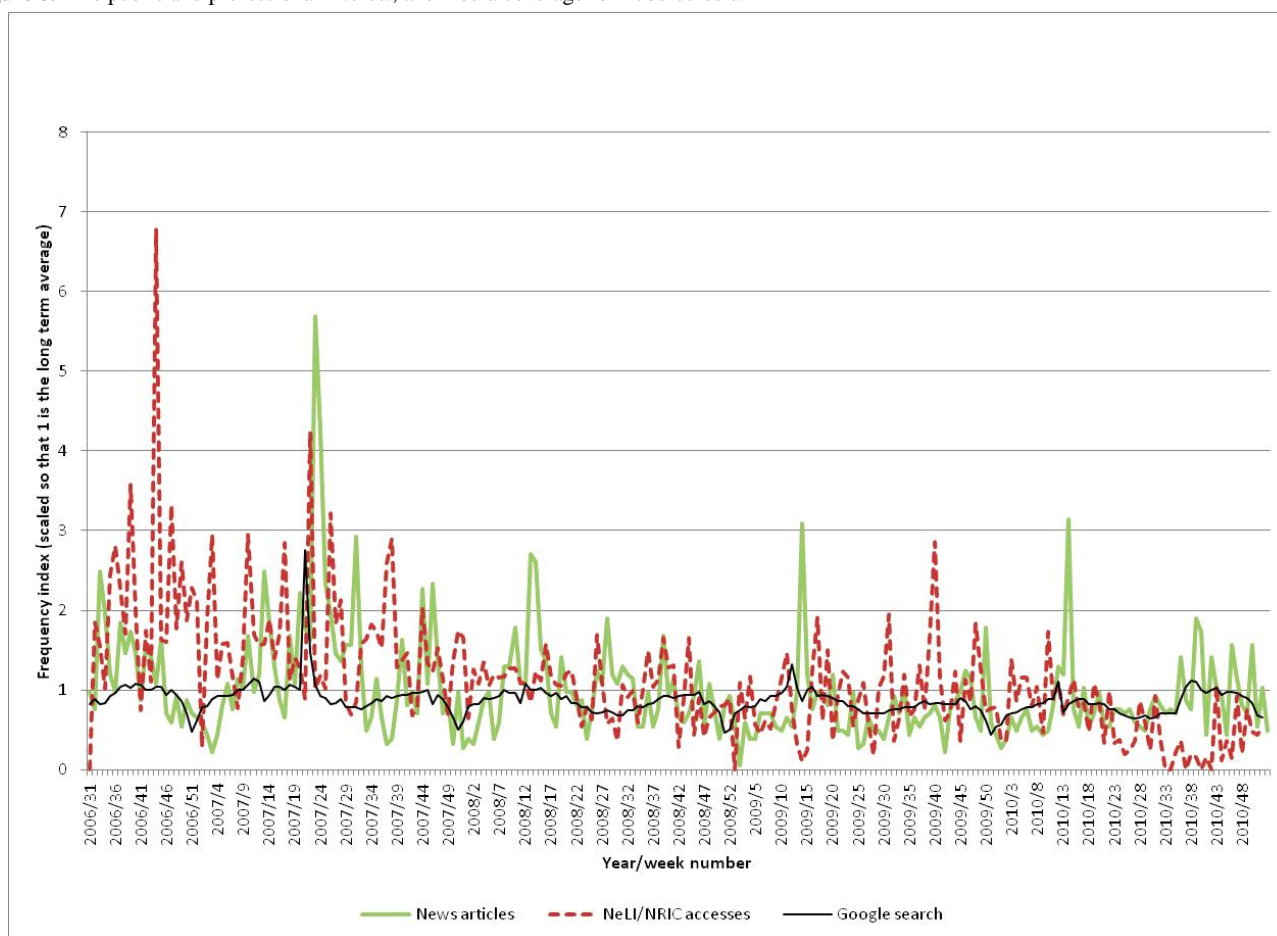
Figure 9. The NeLI/NRIC accesses for TB, and the accesses for the document "Tuberculosis prevention and treatment" published in June 2007.
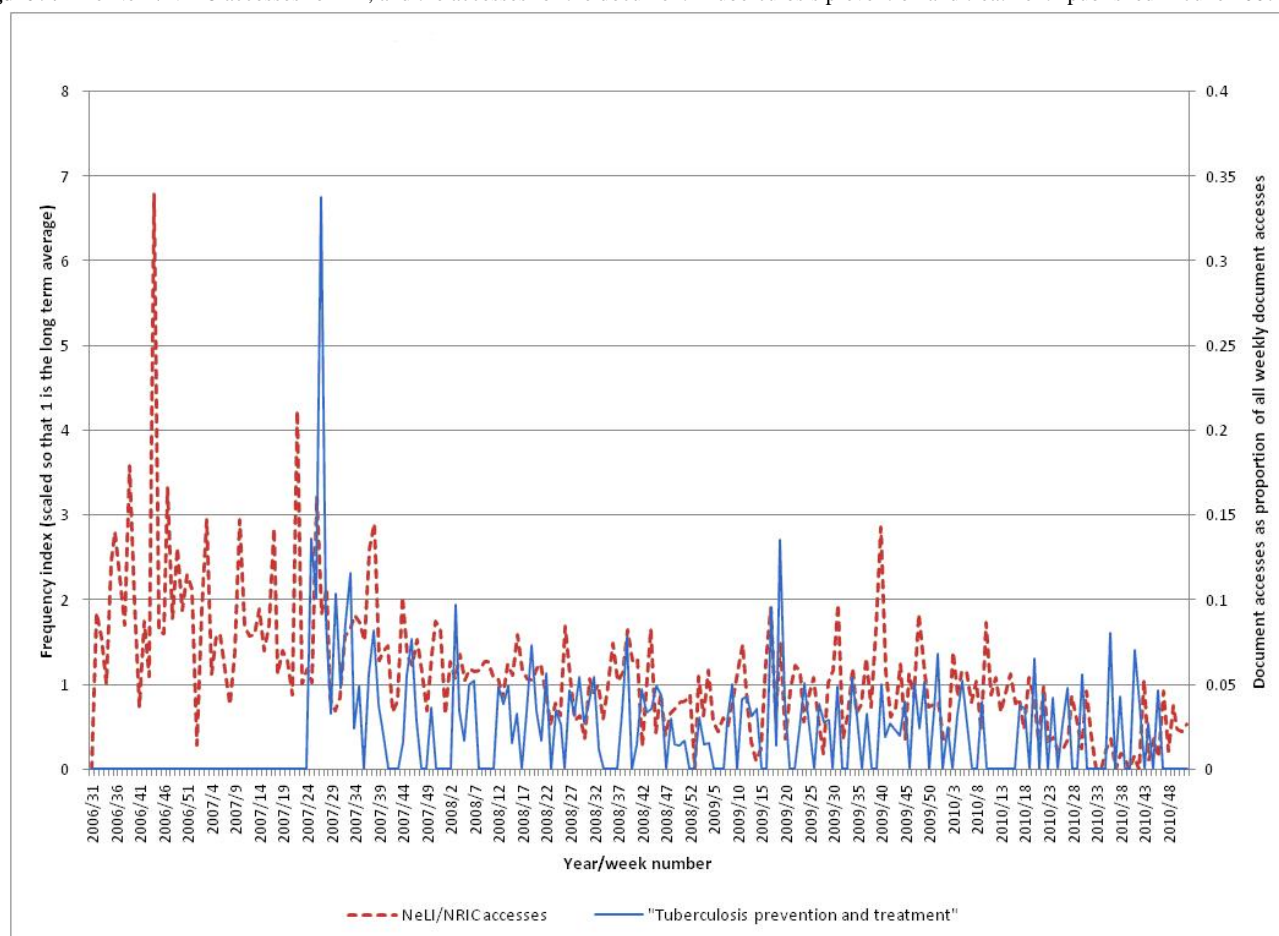
Meningitis

Introduction

Meningitis is "an infection of the meninges (the protective membranes that surround the brain and spinal cord)" and can be caused by either bacteria or viruses [44].

Results

Figure 10 shows the professional interest, public interest, and media coverage for meningitis.

Professional Interest

Professional interest was heightened in the years 2006 and 2007 but has declined since then. Again, some of this decline can be explained as due to enhanced levels of interest when the NeLI/NRIC sites were relaunched in 2006.

Public Interest

The public interest appears to be level, not deviating far from the baseline level, possibly showing a downward trend, as the graph from early 2009 is below the baseline level. The larger dips at the end of 2006 and the end of 2009 are again probably due to the holiday period. This is interesting, as it seems that the presence of heightened media coverage is not influencing the public searches.

Media Coverage

The peaks in media coverage show no obvious pattern. On examining the peaks and matching them to news articles, it seems that media coverage is driven by stories about individual tragedies, celebrity stories, and other human interest stories. The four largest peaks are in week 20, 2007 (3.0 times baseline), with the main focus on the singer Peter Andre who contracted meningitis [45]; in week 39, 2007 (2.5 times baseline), with stories about an individual boy's death [46] and others about a boy who changed his accent after surgery for meningitis [47]; week 12, 2009 (2.5 times baseline), where there was no clear single focus for stories; and week 24, 2009 (2.4 times baseline), focusing on the joint suicide of the parents of a toddler who died of meningitis [48].

Figure 10. The public and professional interest, and media coverage for Meningitis.
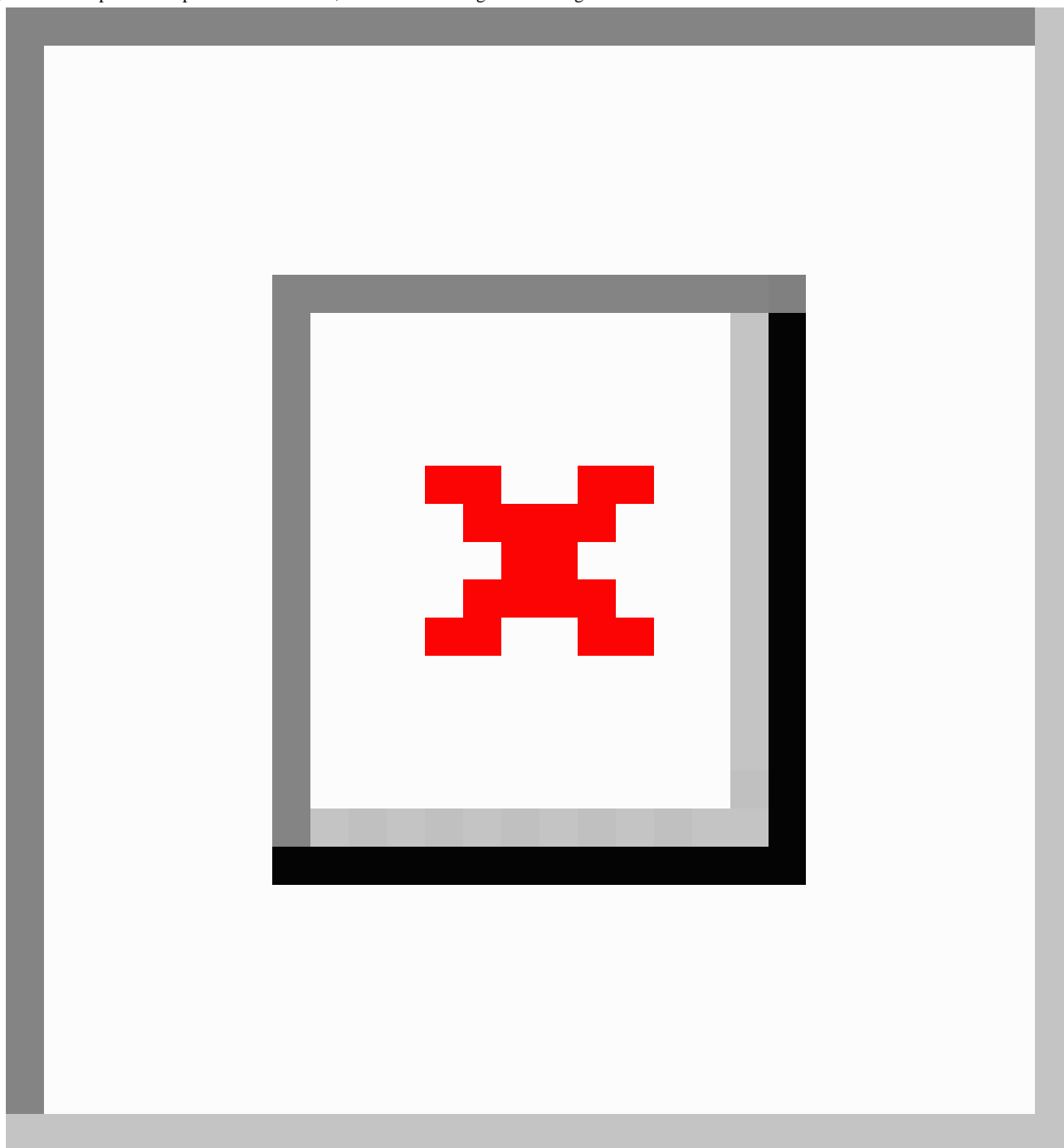
Norovirus

Introduction

Norovirus is a seasonal disease, also called "winter vomiting disease". "The virus, which is highly contagious, causes vomiting and diarrhoea. [...] Between 600,000 and one million people in the UK catch norovirus every year" [49].

Results

In contrast to the other outbreaks that we have investigated in this study, there is a clear seasonal trend to the professional and public interest and media coverage (see Figure 11). Professional interest tends to mirror public interest, except for heightened activity in late 2009 and early 2010. This extra activity is

probably due to publication of the HPA's guideline "Norovirus outbreak reporting scheme" on December 14, 2009.

Figure 12 shows the professional interest (the same NeLI/NRIC accesses as in Figure 11) for norovirus, together with the accesses for the document (measured as a proportion of the total weekly document accesses). There is a small peak in interest at the time of publication (the small size of this peak is due, in part at least, to the document being published near the end of that week), followed by a large peak in the next (full) week, and then a gradual decline.

Professional Interest

Professional interest tends to mirror public interest, except for heightened activity in late 2009 and early 2010 (up to 8.2 times

baseline). This extra activity is due to publication of the HPA's guideline "Norovirus outbreak reporting scheme" in December 2009. The season 2009-2010 was also a "bad" year for norovirus outbreaks (see Figure 13).

Public Interest

The public interest clearly correlates to the seasonal variation in the professional interest graph and in the media coverage. The dips at the end of each year that are clearly visible in the earlier graphs do not appear here.

Media Coverage

There is a clear peak (15.1 times baseline) in media coverage in the winter of 2007/2008. There is also another clear spike (8.6 times baseline) at week 29 of 2009, mainly due to coverage of an outbreak of norovirus on a cruise ship [50]. The media coverage does not fit the seasonal pattern quite as well as the public and professional interest. This may be due to occasional outbreaks that can happen in summer (often on cruise ships)

and also due to the media coverage including a large proportion of stories from the southern hemisphere. The professional interest levels are skewed towards the northern hemisphere, as that is where the majority of visits come from. The public interest levels are also biased towards the northern hemisphere, as the majority of Google's traffic comes from there.

Comparison With Actual Disease Occurrence Data

As the professional interest, public interest, and media coverage show such similar seasonal patterns, it is not surprising that they closely match data for the actual occurrence of the disease. Figure 14 shows the numbers of laboratory reports of norovirus in England and Wales (data from the HPA weekly epidemiological surveillance reports). As the data is from the northern hemisphere it is important to be careful in making generalizations. But there is clearly a close correspondence with the graphs of Figure 11, with the peaks occurring in the northern winter, and the troughs in the summer.

Figure 11. The public and professional interest, and media coverage for Norovirus.
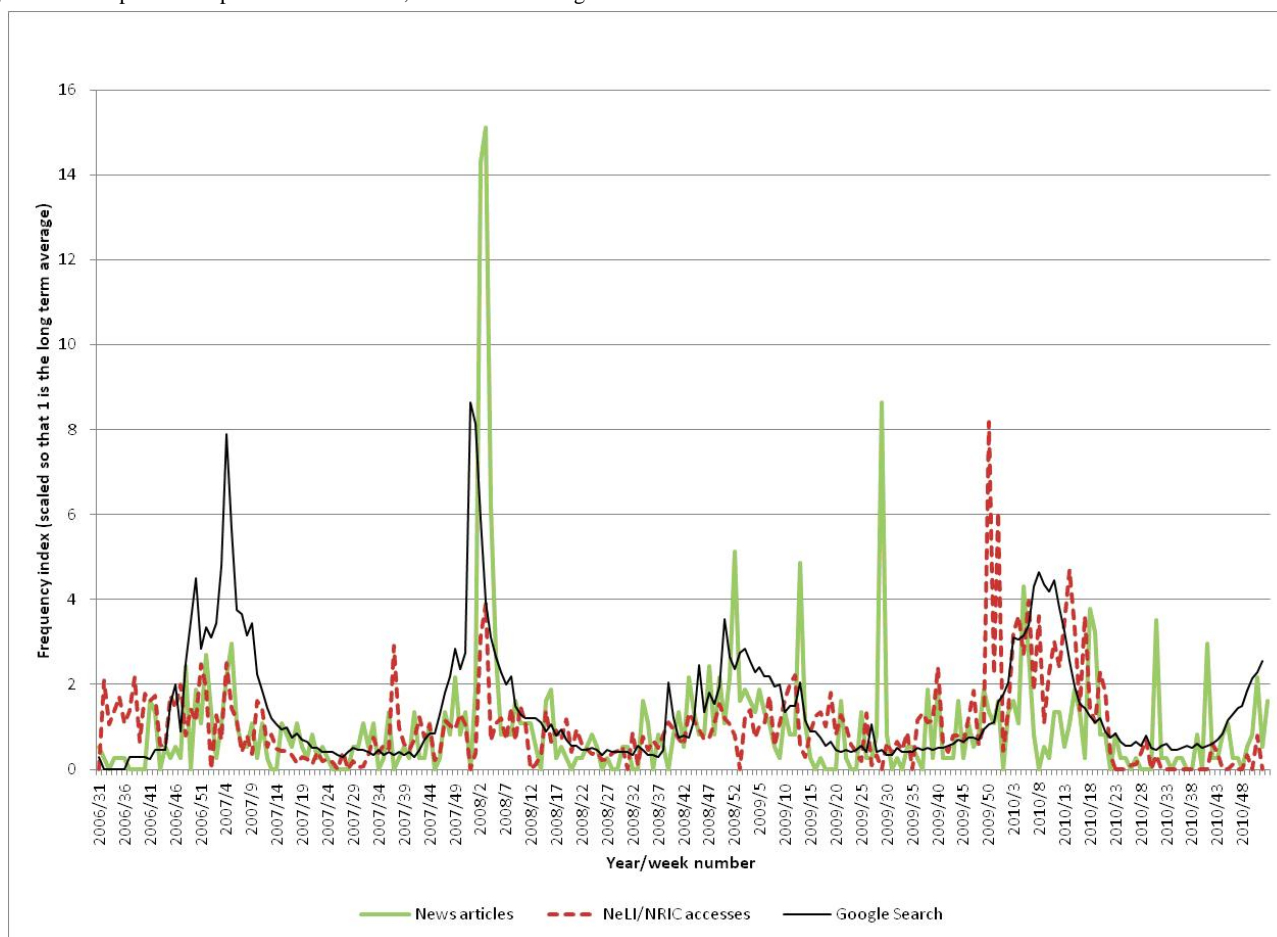
Figure 12. The NeLI/NRIC accesses for norovirus, and the accesses for a specific document published in December 2009.
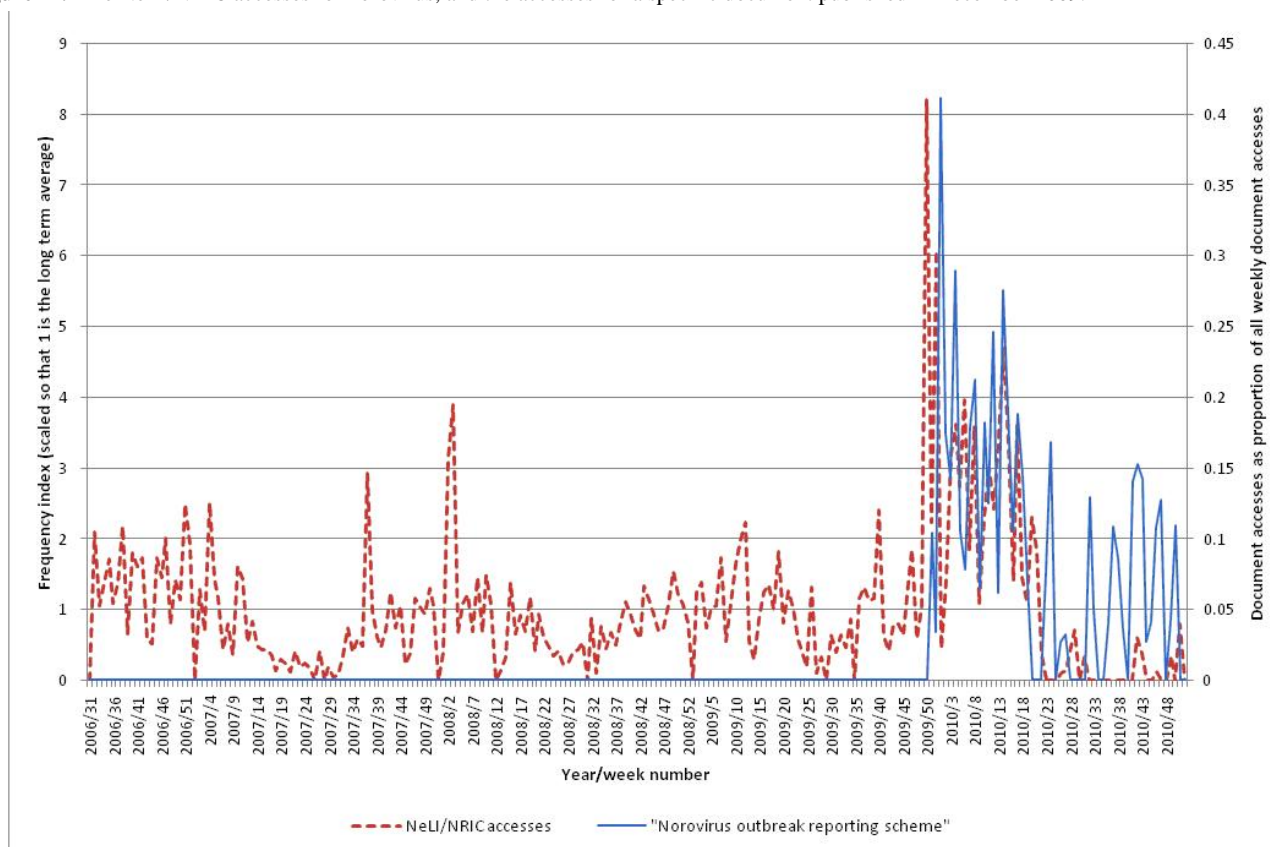
Figure 13. Laboratory reports of norovirus from the years 2005 to 2012 (source: Seasonal comparison of laboratory reports of norovirus (England and Wales; HPA).
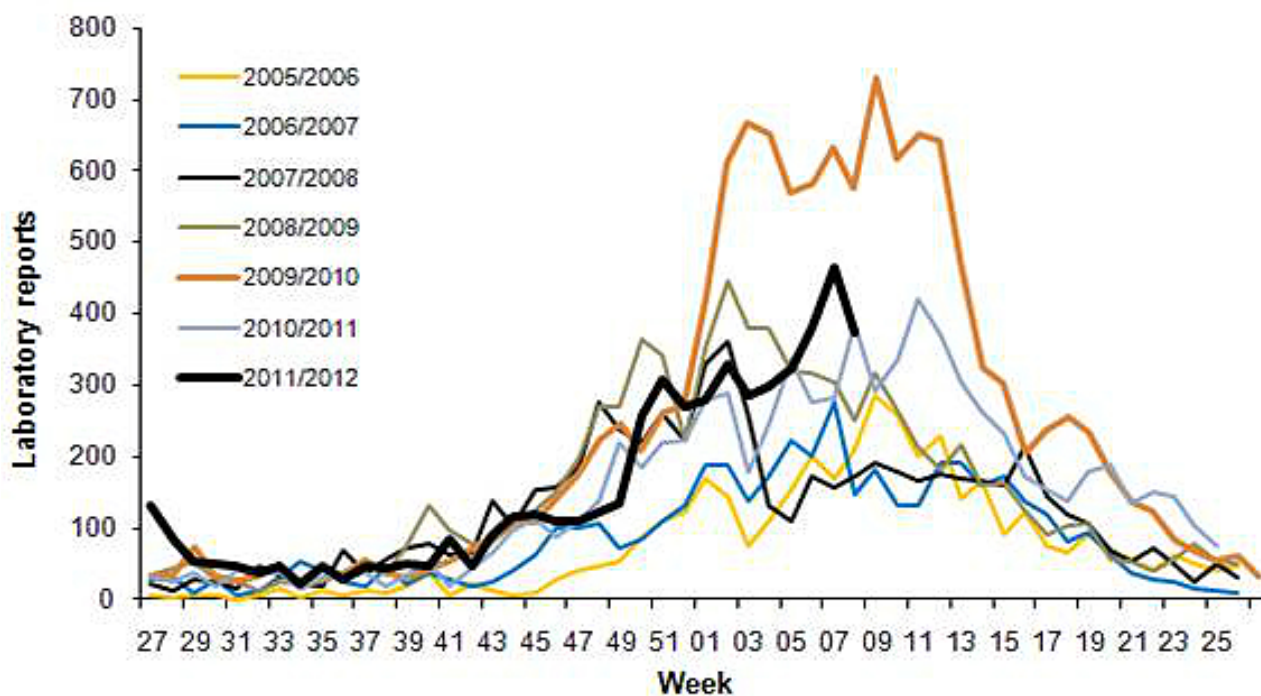
Figure 14. Laboratory reports of norovirus 2004-2011 (England and Wales; graph from HPA).
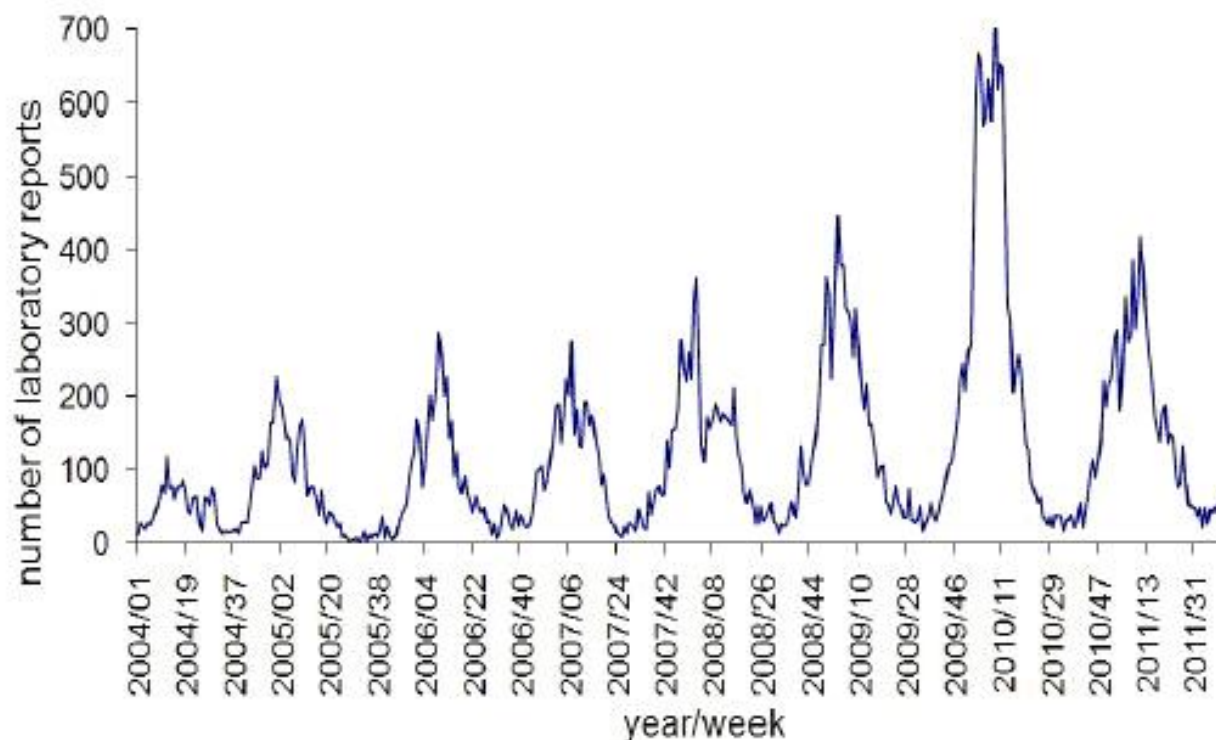
Influenza

Introduction

"Influenza is a viral infection that affects mainly the nose, throat, bronchi and, occasionally, [complications occur which affect the] lungs. [...] The virus is transmitted easily from person to person via droplets and small particles produced when infected people cough or sneeze. Influenza tends to spread rapidly in seasonal epidemics." [51]. New strains and variants of the influenza virus are constantly emerging. There are several named subsets of these, including seasonal influenza, swine influenza ("swine flu"), and avian influenza ("bird flu").

"Seasonal flu occurs every year, usually in the winter. It's a highly infectious disease caused by a virus. The most likely viruses that will cause flu each year are identified in advance and vaccines are then produced that closely match them." [52].

According to an internationally accepted standard, the terms "avian influenza" and "swine influenza" refer to influenza viruses found in birds and swine, respectively [53] .However, the terms (also called "bird flu" and "swine flu") may be used to refer to specific strains of influenza. For example, according to NHS Choices, swine flu is "the common name given to a relatively new strain of influenza (flu) that caused a flu pandemic in 2009-2010. It is also referred to as H1N1 influenza (because it is the H1N1 strain of virus)" [54].

While it may be expected that influenza would feature highest in the NeLI/NRIC accesses over the investigated period, it is actually only in sixth position. The interest in this disease on NeLI/NRIC has shown to be lower as seasonal influenza, being one of the most common diseases, does not require regular and more specialized evidence.

Results

Figure 15 shows the professional interest, public interest, and media coverage for "influenza". Clearly the graph is dominated by the surge of interest around the 2009 swine flu pandemic. Otherwise there was constant public interest in the disease while professionals had several spikes mostly in winter months indicating an increased information need around seasonal influenza. Apart from two isolated media interests in spring 2006 and winter 2007, there were no significant outbreaks resulting in media attention. The key exception requiring an in-depth evaluation is indeed the 2009 swine flu pandemic. Figure 16 shows just the period of 2009 and the first quarter of 2010 (the duration of the swine flu pandemic). Both the public interest and the media coverage have large peaks in the spring of 2009 (36.4 and 12.0 times the baselines, respectively), corresponding to the initial cases in Mexico and the announcement of the pandemic, but have much lower activity later when there was a second peak in flu cases (matching the findings of [17] for the media coverage). However, the public

interest does show a smaller peak (6.4 times baseline) in autumn 2009.

Professional Interest

There is once again a heightened level of activity in the second half of 2006, presumably caused by promotional activity of NeLI/NRIC. During the H1N1 outbreak, the levels were higher than average, but there was not a large spike, as was the case for public interest and media coverage. This is probably due to the large number of competing information online resources for public and professionals that were created during the 2009 pandemic, and the public health agencies such as HPA, ECDC, and WHO held daily press conferences publishing the latest evidence and advice. For this reason, in summer 2009, NeLI/NRIC decided to add a dedicated swine flu link to their home pages to redirect visitors to the ECDC flu website for the daily updates [21].

Public Interest

Interestingly, public interest in the disease peaks in week 17 of 2009, while the peak in media coverage followed in week 18. However, the difference of one week is probably not significant here, as the process of collating results into weekly totals will have some uncertainties. According to a study conducted in 2009:

> The highest number of articles (842) was recorded on 27 April, the day WHO raised the level of influenza pandemic alert to phase 4...There was a smaller, though still large, peak of the number of media articles on 30 April (717 articles). This appears to be linked to WHO's announcement of pandemic alert phase 5 at 22:00 Central European Time on 29 April: many of the European media reports about this were published on 30 April. Media interest dropped considerably after 30 April. [55]

April 27 is near the end of week 17 (23-29 April), and April 30 is at the start of week 18. It is interesting that the announcement of phase 6 (the pandemic level) on June 11, 2009, did not seem to generate any significant interest.

There was a second peak in public interest at week 44 (6.4 times baseline), identifying the autumn 2009 outbreak (this is discussed in the next section, where it is correlated with the media coverage).

Media Coverage

There is an earlier peak (10.0 times baseline) in media coverage in August 2007, which corresponds to a serious outbreak of influenza in Australia (for example [56]).

The second main peak (12.0 times baseline) occurred around the end of April 2009. This was around 6 weeks before the WHO declared that H1N1 was officially a pandemic (June 11, 2009). The heightened media interest in these weeks related to the outbreak in Mexico and speculation as to whether the disease would spread. Public interest during the autumn second peak of the disease was 5.7 times smaller than during the April peak, but is clearly visible in Figure 16, although the media did not give the topic much attention at this crucial time.

Figure 15. The public and professional interest, and media coverage for Influenza.
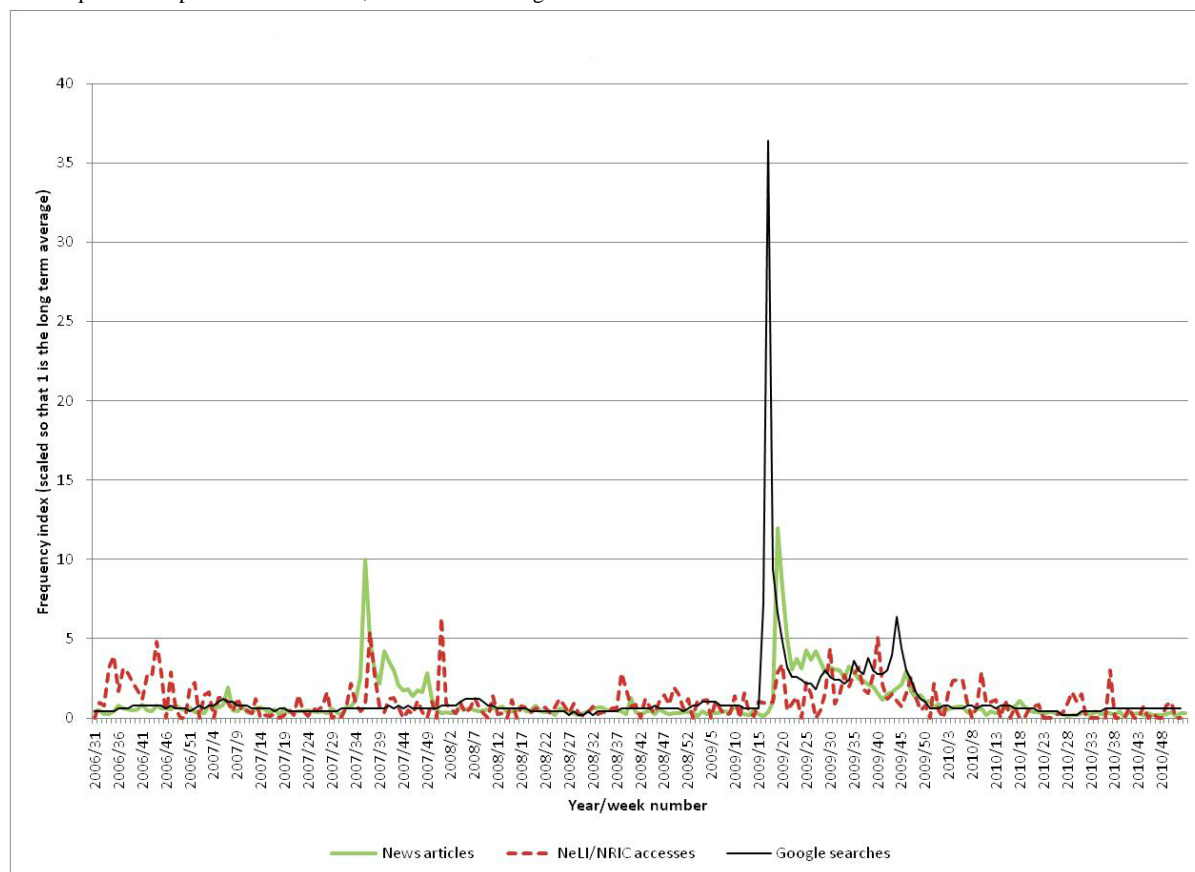
Figure 16. The public and professional interest, and media coverage for Influenza focusing on the period of the 2009-10 swine flu pandemic.
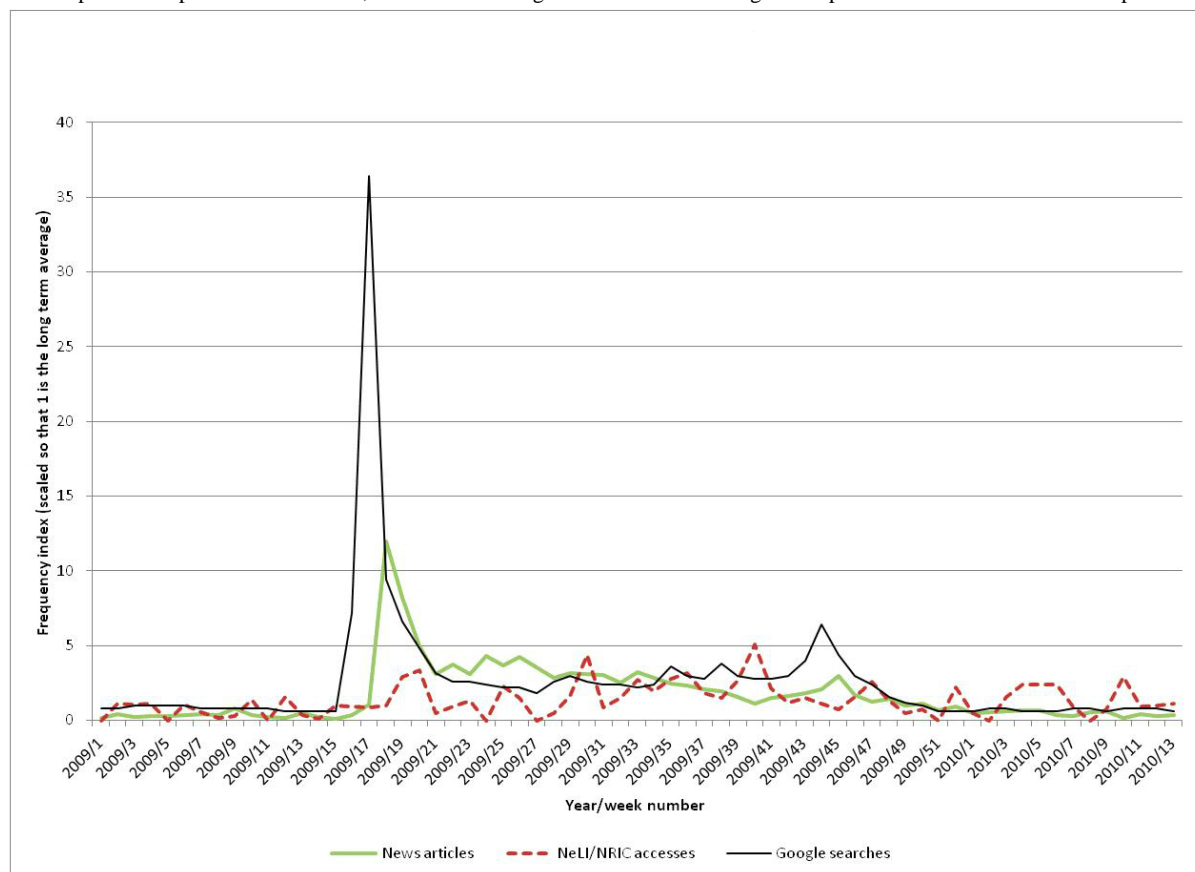
Discussion

Principal Results

In this paper we have analyzed the information needs of public and professionals around key infectious disease outbreaks and events in the 4.5 years from the end of July 2006, until the end of 2010. We compared these with media coverage to illustrate where the media interest could have fueled public interest in the disease and what the reaction was of professionals to key outbreaks and policy changes. Based on the results, the diseases fall into 4 groups:

- MRSA and *Clostridium difficile*: High prevalence, reducing rapidly with new government targets and emphasis on surveillance/reporting;
- Tuberculosis and meningitis: low prevalence;
- Norovirus: seasonal; and
- Influenza: 2009 mass media attention and pandemic event.

The next sections will discuss the results in more detail.

We found that a triangulation of (1) longitudinal Web log data from the NeLI/NRIC infection portals to evaluate the professionals' needs around infection as a primary goal, (2) Google Trends in these topics to find a complementary public interest, and (3) media coverage from LexisNexis provides the desired correlation to answer our research questions listed in section 3:

1. How do health care professionals' online search needs around infection differ from public needs?

2. Does media coverage contribute to the information needs for infection events among public?

3. How are incidents of a disease and major policy events related to information needs of professionals?

Our findings include (corresponding research questions 1-3 are listed in parentheses after each finding):

1. We found that public needs in infection are much more static and do not relate to disease occurrence and media coverage as much as professionals whose needs inevitably increase with a public health event or a key policy change. (for all diseases examined except influenza discussed below) (RQ1).

2. However, for events of major media interest, such as MRSA/*C difficile*, media coverage resulted in a major public interest (such as the late 2007-early 2008 UK outbreak). (RQ2).

3. Meningitis was a clear example of a disease that has a heightened media coverage that tends to focus on individual tragic cases and celebrity stories (RQ2).

4. Professionals' interest did not follow media coverage, but spikes in interest occurred during outbreaks (MRSA, *C difficile*) release of major national policy or important document (for example, the Healthcare Commission report on *C difficile* "Investigation into outbreaks of *C difficile* at Maidstone and Tunbridge Wells NHS Trust" and the HPA document on norovirus in 2009 "Norovirus outbreak reporting scheme") (RQ1).

5. An exception was norovirus, which showed a seasonal pattern for both groups and matched the periodic disease occurrence (RQ3).

6. Influenza was of a major concern during the H1N1 outbreak in 2009, creating massive information needs among the public. Also in autumn 2009, the public interest again peaked, but on a smaller scale and also irrespective of the media coverage. However, the media coverage was on a large scale around June 2009 when WHO officially declared the H1N1 outbreak to be a "phase 5" pandemic (RQ3).

7. Additional results (not corresponding to original research questions RQ1-3): The professional interest was heightened early in the study period for all diseases. This appears to be due to the promotional activities that surrounded the relaunch of NeLI/NRIC in 2006. The professional interest reverted to a more even level after a few months. This is also reflected in the overall graph of traffic for NeLI/NRIC (see Figure 4). Finally, public interest is often difficult to quantify due to the plain text nature of searches and the fact that slang terms are often used (for example "superbug", Figure 6).

In general, we concluded that media plays a role in influencing public information needs but is not as crucial as is often assumed. Professionals naturally respond to disease occurrence, events, or publication of key documents or policy changes that drive their information needs.

Limitations

Studying information online needs is very difficult, and research seems to pay little attention to uncontrolled study and analysis of Web server logs for professional and public information needs. Due to the nature of the data available, we have had to make a number of assumptions in this study:

- We have assumed that a majority of NeLI/NRIC users are health care professionals, compared to Google searches. So we therefore assume that NeLI/NRIC accesses better reflect the interests of health care professionals, whereas Google searches reflect the interests of the wider public. We claim that this assumption is reasonable, as (1) the NeLI/NRIC websites are designed to provide specialist information targeted and promoted at infection professionals, which would be of less interest to the general user, and (2) there are many websites that are more accessible to the general user (such as NHS Choices [25]).

- We have assumed that the number of newspaper articles found via LexisNexis mentioning a keyword near the start of the article is a suitable measure of media coverage. More complex measures could be used, perhaps taking into account the number of words in the article or the readership of the newspapers. Furthermore, other media could be considered, such as television and radio, or social media, such as Twitter.

- We assumed that levels of keyword searches were sufficient to measure interest in particular topics. For NeLI/NRIC, we measured the accesses of a particular topic page, whereas for media coverage and Google searches, we used specific keywords (due to the nature of the available data). In these cases, we are ignoring possible misspellings and synonyms that would have reflected interest in the topic.

- For commercial reasons, Google does not release details of how its search engine algorithms work, and so it is difficult to determine exactly what the Google Trends data represents (whether it includes misspellings and synonyms, for example).

- There are many limitations to using Web server logs to analyze user behavior: (1) it is not possible to resolve IP addresses to individual users as one IP address can represent many users, (2) despite all the efforts discussed we've discussed, it is not possible to identify all non-human users, eg, spiders and crawlers, and importantly (3) Web logs do not provide any insight into *why* users did what they did on the site and whether they were or were not dissatisfied with the results [21].

Conclusions

In the last two decades, the Internet has revolutionized the way we seek up-to-date evidence and information for public, in particular, during major infection events and outbreaks. Also, the role of online media with increasing coverage of public health events has contributed to the demand for information. In this study, we compared professional and public online information needs around major infection events and outbreaks over the period from mid-2006 to the end of 2010, as well as relevant media coverage.

We investigated in depth six diseases with the highest online traffic on NeLI/NRIC: *Clostridium difficile*, MRSA, tuberculosis, meningitis, norovirus, and influenza. The results illustrated that public information needs remain steady and do not necessarily follow media coverage unless the event is widely covered (MRSA/*C difficile* and influenza).

As expected, professionals' interest did not follow media coverage but spikes in interest occurred during major outbreaks (MRSA and *C difficile*) and around the release of major national policy or other important documents (eg, the Healthcare Commission's report on *C difficile*, entitled "Investigation into outbreaks of *Clostridium difficile* at Maidstone and Tunbridge Wells NHS Trust") and the HPA document on norovirus in 2009 ("Norovirus outbreak reporting scheme"). The exception was norovirus showing a seasonal pattern for both groups and matching the periodic disease occurrence. Influenza was of a major concern during the H1N1 outbreak in 2009 creating massive information needs among the public in the spring in line with the media coverage and again in the autumn of 2009, this time regardless of the media coverage.

Therefore, public health agencies with responsibility for risk communication of public health events, in particular during outbreaks and emergencies, need to collaborate with media in order to ensure the coverage is of the highest quality and evidence-based while professionals information needs remain mainly fulfilled by online open access to key resources.

Acknowledgments

Conflicts of Interest

None declared.

Multimedia Appendix 1

Supplementary figures.

[PDF File (Adobe PDF File), 356KB - jmir_v15i7e107_app1.pdf]

References

1. Higgins O, Sixsmith J, Barry MM, Domegan C. A literature review on health information-seeking behavior on the Web: a health consumer and health professional perspective, Stockholm.: European Centre for Disease Control and Prevention; 2011. URL: http://ecdc.europa.eu/en/publications/Publications/Literature%20review%20on%20health%20information-seeking%20behaviour%20on%20the%20web.pdf [accessed 2013-04-26] [WebCite Cache ID 6GAOy0LcS]
2. Coote A, Harrabin R, Allen J. Health in the News: Risk, Reporting and Media Influence. London: King's Fund; 2003.
3. Seybert H. Internet use in households and by individuals in 2011. Eurostat. Statistics in Focus 2011;66.
4. Fox S, Jones S. The social life of health information.: Pew Internet and American Life Project; 2009. URL: http://www.pewinternet.org/Reports/2009/8-The-Social-Life-of-Health-Information.aspx [accessed 2013-04-26] [WebCite Cache ID 6GAPPTJ3U]
5. Taylor H. Healthcare News. Harris Interactive 2008:8.
6. Andreassen K. European citizens' use of E-health services: a study of seven countries. BMC public health 2007;7:53. [doi: 10.1186/1471-2458-7-53]
7. Baker L, Wagner TH, Singer S, Bundorf MK. Use of the Internet and e-mail for health care information: results from a national survey. JAMA 2003 May 14;289(18):2400-2406. [doi: 10.1001/jama.289.18.2400] [Medline: 12746364]
8. Fox S. Mobile Health 2010.: Pew Internet and American Life Project; 2010. URL: http://pewinternet.org/Reports/2010/Mobile-Health-2010.aspx [accessed 2013-04-26] [WebCite Cache ID 6GAPn2xUa]
9. Williams P, Madle G, Weinberg J, Kostkova P, Mani-Saada J. Information for the public about disease: usability issues in the development of the National Electronic Library for Communicable Diseases. Aslib Proceedings 2004;56(2):99-103. [doi: 10.1108/00012530410529468]
10. Younger P. Internet-based information-seeking behaviour amongst doctors and nurses: a short review of the literature. Health information and libraries journal 2010 Mar;27(1):2-10.
11. Adams A, Blandford A, Attfield S. Implementing digital resources for clinicians' and patients' varying needs. Med Inform Internet Med 2005 Jun;30(2):107-122. [doi: 10.1080/14639230500298875] [Medline: 16338799]
12. Alghanim S. Information needs and seeking behavior among primary care physicians in Saudi Arabia?: Implications for policy and practice. Scientific Research and Essays 2011;6(8):1849-1855.
13. O'Keeffe J, Willinsky J, Maggio L. Public access and use of health research: an exploratory study of the National Institutes of Health (NIH) Public Access Policy using interviews and surveys of health personnel. J Med Internet Res 2011;13(4):e97 [FREE Full text] [doi: 10.2196/jmir.1827] [Medline: 22106169]
14. Greenberg M, Wartenberg D. Understanding mass media coverage of disease clusters. Am J Epidemiol 1990 Jul;132(1 Suppl):S192-S195. [Medline: 2356831]
15. Godlee F, Smith J, Marcovitch H. Wakefield's article linking MMR vaccine and autism was fraudulent. BMJ 2011 Jan 2011;342(jan05 1):c7452-c7452. [doi: 10.1136/bmj.c7452]
16. Mummert A, Weiss H. Get the News Out Loudly and Quickly: Modeling the Influence of the Media on Limiting Infectious Disease Outbreaks. Quantitative Biology 2010 Jun 25 [FREE Full text]
17. Hilton S, Hunt K. UK newspapers' representations of the 2009-10 outbreak of swine flu: one health scare not over-hyped by the media? J Epidemiol Community Health 2011 Oct;65(10):941-946 [FREE Full text] [doi: 10.1136/jech.2010.119875] [Medline: 21131303]

18. Seale H, Heywood AE, McLaws ML, Ward KF, Lowbridge CP, Van D, et al. Why do I need it? I am not at risk! Public perceptions towards the pandemic (H1N1) 2009 vaccine. BMC Infect Dis 2010;10:99 [FREE Full text] [doi: 10.1186/1471-2334-10-99] [Medline: 20403201]

19. Ammenwerth E, Iller C, Mansmann U. Can evaluation studies benefit from triangulation? A case study. Int J Med Inform 2003 Jul;70(2-3):237-248. [Medline: 12909175]

20. Google. Google Trends. URL: http://www.google.com/trends [accessed 2012-01-04] [WebCite Cache ID 64RSbbPmq]

21. Madle G, Kostkova P, Mani-Saada J, Roy A. Lessons learned from evaluation of the use of the National electronic Library of Infection. Health Informatics J 2006 Jun;12(2):137-151. [doi: 10.1177/1460458206063809] [Medline: 17023404]

22. NeLI. National electronic Library of Infection (NeLI). URL: http://www.neli.org.uk/IntegratedCRD.nsf/NeLI_Home1?OpenForm [accessed 2012-01-04] [WebCite Cache ID 64RSjNp3L]

23. Roy A, Kostkova P, Catchpole M, Carson E. Web-based provision of information on infectious diseases: a systems study. Health Informatics J 2006 Dec;12(4):274-292. [doi: 10.1177/1460458206069760] [Medline: 17092999]

24. NRIC. National Resource for Infection Control (NRIC). URL: http://www.nric.org.uk/IntegratedCRD.nsf/NRIC_Home1?OpenForm [accessed 2012-01-04] [WebCite Cache ID 64RStS37P]

25. National Health Service. NHS Choices. URL: http://www.nhs.uk/Pages/HomePage.aspx [accessed 2012-03-28] [WebCite Cache ID 66UyCjtEM]

26. City eHealth Research Centre. Bugs and Drugs on the Web. URL: http://www.antibioticresistance.org.uk/ [accessed 2012-03-28] [WebCite Cache ID 66UyNl6nd]

27. Fox S, Rainie L. Vital Decisions: A Pew Internet Health Report.: Pew Internet and American Life Project; 2002. URL: http://www.pewinternet.org/Reports/2002/Vital-Decisions-A-Pew-Internet-Health-Report.aspx [accessed 2013-04-26] [WebCite Cache ID 6GARMJyyY]

28. Eysenbach G, Köhler C. How do consumers search for and appraise health information on the world wide web? Qualitative study using focus groups, usability tests, and in-depth interviews. British Medical Journal 2002 Mar 9;324:573.

29. StatCounter. StatCounter Global Stats. URL: http://gs.statcounter.com/ [accessed 2012-03-28] [WebCite Cache ID 66UxiYpfw]

30. LexisNexis. LexisNexis. URL: http://w3.nexis.com/new/ [accessed 2012-03-22] [WebCite Cache ID 66Ls0ZcNC]

31. Jawaheer G, Kostkova P. Web crawlers on a health related portal: detection, characterisation and implications. In: Proceedings of the 2011 Developments in E-systems Engineering.: IEEE; 2011 Presented at: 2010 Developments in E-systems Engineering; Sept. 6-8, 2010; London, UK p. 24-29 URL: http://ieeexplore.ieee.org/xpl/login.jsp?tp=&arnumber=6149949&url=http%3A%2F%2Fieeexplore.ieee.org%2Fxpls%2Fabs_all.jsp%3Farnumber%3D6149949 [doi: 10.1109/DeSE.2011.83]

32. Huntington P, Nicholas D, Jamali H. Web robot detection in the scholarly information environment. Journal of Information Science 2008 Apr;34(5):726-741.

33. Wired. 2002. When the Spam Hits the Blogs URL: http://www.wired.com/culture/lifestyle/news/2002/10/56017 [accessed 2013-04-26] [WebCite Cache ID 6GASCaSHY]

34. Roy A, Kostkova P, Catchpole M, Carson E. 'Do Users Do What They Think They Do?'-A Comparative Study of User Perceived and Actual Information Searching Behaviour in the National Electronic Library of Infection. Electronic Healthcare 2010;27:96-103. [doi: 10.1007/978-3-642-11745-9_15]

35. National Health Service. Clostridium difficile - NHS Choices. 2010. URL: http://www.nhs.uk/conditions/clostridium-difficile/pages/introduction.aspx [accessed 2012-03-09] [WebCite Cache ID 662IGs06L]

36. National Health Service. MRSA infection - NHS Choices. 2010. URL: http://www.nhs.uk/conditions/mrsa/pages/introduction.aspx [accessed 2012-03-09] [WebCite Cache ID 662Iq9ou8]

37. Google. Google Adwords. URL: http://www.google.com/adwords/?sourceid=awo&subid=ww-et-awhp_nelsontest3_nel_p&clickid [WebCite Cache ID 64vpo2lGc]

38. Carvel J. The Guardian. 2007. NHS superbug inquiry blames trust for deaths URL: http://www.guardian.co.uk/society/2007/oct/11/health?INTCMP=SRCH [accessed 2012-01-04] [WebCite Cache ID 64RSSYI1h]

39. Lite J. New York Daily News. 2007 May 30. TB case spurs panic URL: http://www.nydailynews.com/news/world/tb-case-spurs-panic-article-1.254526 [accessed 2013-04-26] [WebCite Cache ID 6GATFMR0F]

40. Altman L, Schwartz J. New York Times. 2007 May 31. Near Misses Allowed Man With Tuberculosis to Fly URL: http://www.nytimes.com/2007/05/31/us/31tb.html?pagewanted=all&_r=0

41. Morris S. The Guardian. 2007. Death sentence confirmed on Hindu bullock URL: http://www.guardian.co.uk/uk/2007/jul/04/religion.animalwelfare [accessed 2012-01-25] [WebCite Cache ID 64x82qzAU]

42. Meikle J. The Guardian. 2008. TB: Farmers ready for fight unless badgers culled URL: http://www.guardian.co.uk/environment/2008/jul/05/wildlife.conservation [accessed 2012-01-25] [WebCite Cache ID 64x8aB5Gv]

43. Boseley S. The Guardian. 2008. Transplant first a giant leap for surgery URL: http://www.guardian.co.uk/society/2008/nov/19/stem-cell-transplant-claudio-castillo [accessed 2012-01-25] [WebCite Cache ID 64x8jCeJC]

44. National Health Service. Meningitis - NHS Choices. 2010. URL: http://www.nhs.uk/conditions/meningitis/pages/introduction.aspx [accessed 2012-01-04] [WebCite Cache ID 64RRdEiaQ]

45. BBC. Singer Andre has meningitis tests. 2007. URL: http://news.bbc.co.uk/2/hi/uk_news/england/surrey/6599671.stm [accessed 2012-04-23] [WebCite Cache ID 678vq3pIh]

46. The Scotsman. 2007 Sep 21. NHS 24 system changed after baby's death URL: http://www.scotsman.com/news/health/nhs-24-system-changed-after-baby-s-death-1-1332057 [accessed 2013-04-26] [WebCite Cache ID 6GAU536VQ]

47. Hull L. Mail Online. 2007 Sep 18. The accent transplant: Brain surgery leaves Yorkshire boy speaking like the Queen URL: http://www.dailymail.co.uk/news/article-482256/The-accent-transplant-Brain-surgery-leaves-Yorkshire-boy-speaking-like-Queen.html [accessed 2013-04-26] [WebCite Cache ID 6GAUHbpPV]

48. Morris S. The Guardian. 2009 Jun 2. Couple made suicide leap with body of son after his death from meningitis URL: http://www.guardian.co.uk/society/2009/jun/02/couple-suicide-son-beachy-head [accessed 2013-04-26] [WebCite Cache ID 6GAUYoxx6]

49. National Health Service. Norovirus Infections - NHS Choices. URL: http://www.nhs.uk/conditions/norovirus/pages/introduction.aspx [accessed 2012-01-04] [WebCite Cache ID 64RRRXeM1]

50. Cramb A. The Telegraph, London. 2009 Jul 08. Vomiting bug ship to set sail for home URL: http://www.telegraph.co.uk/news/uknews/scotland/5779481/Vomiting-bug-ship-to-set-sail-for-home.html [accessed 2013-04-26] [WebCite Cache ID 6GAUlDK5K]

51. World Health Organisation. Health Topics - Influenza. 2012. URL: http://www.who.int/topics/influenza/en/ [accessed 2012-03-22] [WebCite Cache ID 66M5J5HWW]

52. Department of Health, Richmond House. Seasonal flu vaccination: Who should have it and why. 2010. URL: https://www.gov.uk/government/publications/seasonal-flu-vaccination-who-should-have-it-and-why-2010 [accessed 2013-04-26] [WebCite Cache ID 6GAVCQCFe]

53. WHO. A revision of the system of nomenclature for influenza viruses: a WHO memorandum. Bull World Health Organ 1980;58(4):585-591 [FREE Full text] [Medline: 6969132]

54. National Health Service. Swine flu (H1N1). 2011. URL: http://www.nhs.uk/Conditions/Pandemic-flu/Pages/Introduction.aspx [accessed 2012-03-22] [WebCite Cache ID 66M8xCIdw]

55. Duncan B. How the media reported the first days of the pandemic (H1N1) 2009: results of EU-wide media analysis. Euro Surveill 2009 Jul 30;14(30):19286 [FREE Full text] [Medline: 19643058]

56. Pirani C. The Australian. 2007 Aug 11. Influenza season could be worst in years URL: http://www.theaustralian.com.au/news/health-science/influenza-season-could-be-worst-in-years/story-e6frg8y6-1111114159692 [accessed 2013-04-26] [WebCite Cache ID 6GAVc4Bg7]

Abbreviations

C difficile: *Clostridium difficile*
CeRC: City eHealth Research Centre
ECDC: European Centre for Disease Prevention and Control
EU: European Union
HAI: health care associated infections
HPA: Health Protection Agency
MMR: measles, mumps, and rubella
MRSA: multi-resistant Staphylococcus aureus
NeLI: National electronic Library of Infection
NHS: National Health Service (UK)
NRIC: National Resource for Infection Control
SARS: Severe Acute Respiratory Syndrome
TB: tuberculosis
WHO: World Health Organization

XSL•FO

RenderX

<u>Original Paper</u>

User Evaluation of the Effects of a Text Simplification Algorithm Using Term Familiarity on Perception, Understanding, Learning, and Information Retention

Gondy Leroy[1,2*], MS, PhD; James E Endicott[1*]; David Kauchak[3*], PhD; Obay Mouradi[1*]; Melissa Just[4*]

[1]Information Systems and Technology, Claremont Graduate University, Claremont, CA, United States

[2]Eller College of Management, Department of Management Information System, University of Arizona, Tucson, AZ, United States

[3]Computer Science Department, Middlebury College, Middlebury, VT, United States

[4]Rutgers University Libraries, Rutgers, The State University of New Jersey, New Brunswick, NJ, United States

[*]all authors contributed equally

Corresponding Author:
Gondy Leroy, MS, PhD
Information Systems and Technology
Claremont Graduate University
ACB 225
130 E Ninth Street
Claremont, CA, 91711
United States
Phone: 1 909 607 3270
Fax: 1 909 621 8564
Email: gondy.leroy@cgu.edu

Abstract

Background: Adequate health literacy is important for people to maintain good health and manage diseases and injuries. Educational text, either retrieved from the Internet or provided by a doctor's office, is a popular method to communicate health-related information. Unfortunately, it is difficult to write text that is easy to understand, and existing approaches, mostly the application of readability formulas, have not convincingly been shown to reduce the difficulty of text.

Objective: To develop an evidence-based writer support tool to improve perceived and actual text difficulty. To this end, we are developing and testing algorithms that automatically identify difficult sections in text and provide appropriate, easier alternatives; algorithms that effectively reduce text difficulty will be included in the support tool. This work describes the user evaluation with an independent writer of an automated simplification algorithm using term familiarity.

Methods: Term familiarity indicates how easy words are for readers and is estimated using term frequencies in the Google Web Corpus. Unfamiliar words are algorithmically identified and tagged for potential replacement. Easier alternatives consisting of synonyms, hypernyms, definitions, and semantic types are extracted from WordNet, the Unified Medical Language System (UMLS), and Wiktionary and ranked for a writer to choose from to simplify the text. We conducted a controlled user study with a representative writer who used our simplification algorithm to simplify texts. We tested the impact with representative consumers. The key independent variable of our study is lexical simplification, and we measured its effect on both perceived and actual text difficulty. Participants were recruited from Amazon's Mechanical Turk website. Perceived difficulty was measured with 1 metric, a 5-point Likert scale. Actual difficulty was measured with 3 metrics: 5 multiple-choice questions alongside each text to measure understanding, 7 multiple-choice questions without the text for learning, and 2 free recall questions for information retention.

Results: Ninety-nine participants completed the study. We found strong beneficial effects on both perceived and actual difficulty. After simplification, the text was perceived as simpler ($P<.001$) with simplified text scoring 2.3 and original text 3.2 on the 5-point Likert scale (score 1: easiest). It also led to better understanding of the text ($P<.001$) with 11% more correct answers with simplified text (63% correct) compared to the original (52% correct). There was more learning with 18% more correct answers after reading simplified text compared to 9% more correct answers after reading the original text ($P=.003$). There was no significant effect on free recall.

Conclusions: Term familiarity is a valuable feature in simplifying text. Although the topic of the text influences the effect size, the results were convincing and consistent.

XSL•FO
RenderX

Introduction

Background and Significance

Text is an important source for health-related information. It is easy to create, maintain, and distribute, and medical practitioners often use it to provide instructions and details on treatments. Health-related text is becoming increasingly available with an estimated 80% of online users [1] from a wide array of backgrounds [2] using the Internet to obtain health-related information. The information itself is diverse and includes prevention, treatment, and management of diseases and comes from a variety of sources ranging from professionals to salespeople to patients.

Unfortunately, 90 million Americans have difficulty understanding and acting upon health information [3], and many find the text currently available difficult to read [4]. Some of this difficulty can be attributed to inherent complexity in understanding the diseases, their causes, and the associated treatments, which may require advanced knowledge of biology, chemistry, or physiology to understand in detail. Much of the difficulty, though, can be attributed to a mismatch between the content delivered and the consumers who often have limited health literacy, low general education, or inadequate language skills. Low health literacy reduces health statuses of individuals [3], is considered a "silent killer" [5], and is estimated to cost up to US$238 billion annually [6].

To increase health literacy, the method, medium, and language used play an important role. While one-on-one teaching may be the best solution, medical professionals do not have sufficient time or resources for this. Video and interactive methods can be very educative and are becoming increasingly available. The power of such methods to teach and demonstrate will likely play an important role in consumer health information. However, currently text remains the primary tool used to educate people.

Factors Influencing Text Difficulty and Its Measurement

Figure 1 provides an overview of three key factors representing the authors' view on influences on understanding and learning from text: personal characteristics, text characteristics, and measurement characteristics. Personal characteristics describe attributes about the reader. Some are innate and cannot be changed, for example, native language and general intelligence. Others are acquired, for example, vocabulary size and domain knowledge. Many of these characteristics have a direct effect on text comprehension and indirectly on learning since comprehension has been shown to affect learning [7]. For example, stress, a personal characteristic, has been shown to affect reading behaviors. People with high stress rely more on visual summaries, even when incomplete, to answer text-based questions [8]. Moreover, increased stress has also been related to lower comprehension of medical terminology [9]. Other personal characteristics, such as the ability to form a good

mental model, affects understanding since readers often rely on the mental model instead of the original text base [10]. In addition, past behaviors and acquired skills can have an impact. Exposure to print, for example, has been found to be related to understanding. Landi [11] found a positive relation with results for an author recognition test [12] and question-answering tasks, while in our own work, we found a positive relation between self-reported reading and results for a fill-in-the-blank Cloze test [13].

Text characteristics influence text difficulty and therefore understanding. These characteristics can be adjusted to improve the usefulness of text, but this has been shown to be challenging and very few studies have shown strong improvements in reader understanding. To further clarify the analysis of text characteristics and the text simplification problem in general, we distinguish between the perceived and actual text difficulty of a text. The distinction is based on evidence for the existence of perceived barriers from the Health Belief Model [14] and the importance of perceived difficulty of behavioral control from the Theory of Planned Behavior [15]. While actual difficulty is easily accepted as important, perceived difficulty cannot be ignored. At a minimum, it will impact whether or not a text will be read. However, it may affect health literacy in more ways; for example, Velayo [16] found that a higher perceived difficulty correlated with a decrease in the recall of information.

Text characteristics can include surface features, for example, spacing or font, and range from smaller units such as words, to larger units such as sentences or paragraphs. Using a Likert scale to measure perceived difficulty, it was found that texts with a higher ratio of function words, verbs, verb phrases, or containing more high-frequency words were seen as easier [17-19]. For actual difficulty, simple surface features such as font and line spacing were shown not to affect remembering [20]; however, using a fill-in-the-blanks test additive and causal connectors were shown to be easier than adversative or sequential connectors [21]. In addition to surface features, analysis can include broader features such as coherence, which is defined as good flow combined with a structured, logical argument [22,23]. We found that increasing coherence with proper spacing around subtopics and better logical connectors improved question-answering (actual difficulty) [13]. Not surprisingly, how a topic is presented in a text also influences learning; topics introduced as part of a refutation text, a text where misconceptions are explicitly addressed, led to increased learning and more valid inference but not increased quantity of information being recalled [7].

Measurement characteristics also play an important role in readability research, although they are often ignored. Historically, the most popular measurement has been readability formulas, which generate a single number often based only on relative word and sentence length and are used as stand-ins for text complexity [24]. These formulas have become popular even though they ignore current knowledge about the reading process,

have a shaky statistical basis, and are unhelpful as writing guidelines [25]. The Flesch-Kincaid Grade Level formula is the most common in health care literature [26]. Even though different tools using the formula sometimes return different levels for the same text [27], it has been used to evaluate patient education materials [28], general websites [29], and information on specific topics such as abdominal aortic aneurysms [30] and back pain [31]. Other readability formulas, such as the Simple Measure of Gobbledygook (SMOG) and Gunning Fog Index, have also been shown to be problematic for evaluating health-related materials for similar reasons [32]. Simplifying text based on these formulas sometimes results in more difficult text, that is, the simplicity paradox [5], because the simplification concentrates on writing style rather than content [2]. As a result, increasingly more concerns are raised about the effectiveness of these formulas for simplifying consumer health texts [33].

Better measures should be developed and used to evaluate text and motivate algorithmic components. These must be evaluated on a representative sample and measure not just the perception of difficulty, but more importantly understanding and retention of information. By using different measures, we can better evaluate the impact of simplification tools. For example, question-answering tasks (eg, multiple-choice, open-ended, or free recall questions), fill-in-the blanks tasks (eg, multiple-choice or open blanks) and teach-back methods (eg, explain a concept or summarize a topic) can be used to measure understanding. Measures that test retention of information can follow the same style of questions, while measures of learning from a text require a comparison between pre- and post-reading scores.

Interactions can also exist between personal, text, and even measurement characteristics. For example, the impact of text coherence on the reader has been found to interact with user characteristics and with the type of measurement. Overall coherence did not affect recall (actual difficulty) but affected remembering and understanding when measured by question-answering (actual difficulty) for readers with high knowledge but low interest or low knowledge but high interest in a topic [10,22]. Personal interest in the topic has also repeatedly been shown to be relevant. A higher interest leads to increased learning [34] and recall [35], however, the coherence of text [34] and prior knowledge [35] influence this relationship.

Objective

Our objective is twofold. First, we address the need for an evidence-based algorithm that pinpoints difficult text. Second, we focus on providing appropriate, easier alternatives to a writer in an effective and efficient manner. We present here our first fully automated version of the *lexical* simplification algorithm, which identifies difficult terms and generates a list of easier alternatives based on information extracted from dictionaries and other databases. In a pilot study [36], we introduced the text simplification algorithm and presented an initial user study. This work builds upon the lessons learned in the pilot study and differs in a number of key dimensions: (1) the algorithm examined here is fully automated, (2) the simplification of text is done by an independent writer, not the developers, and (3) the evaluation is based on a new study with different participants, new stimuli, and new more comprehensive metrics.

Figure 1. Factors that influence understanding and retention of information.
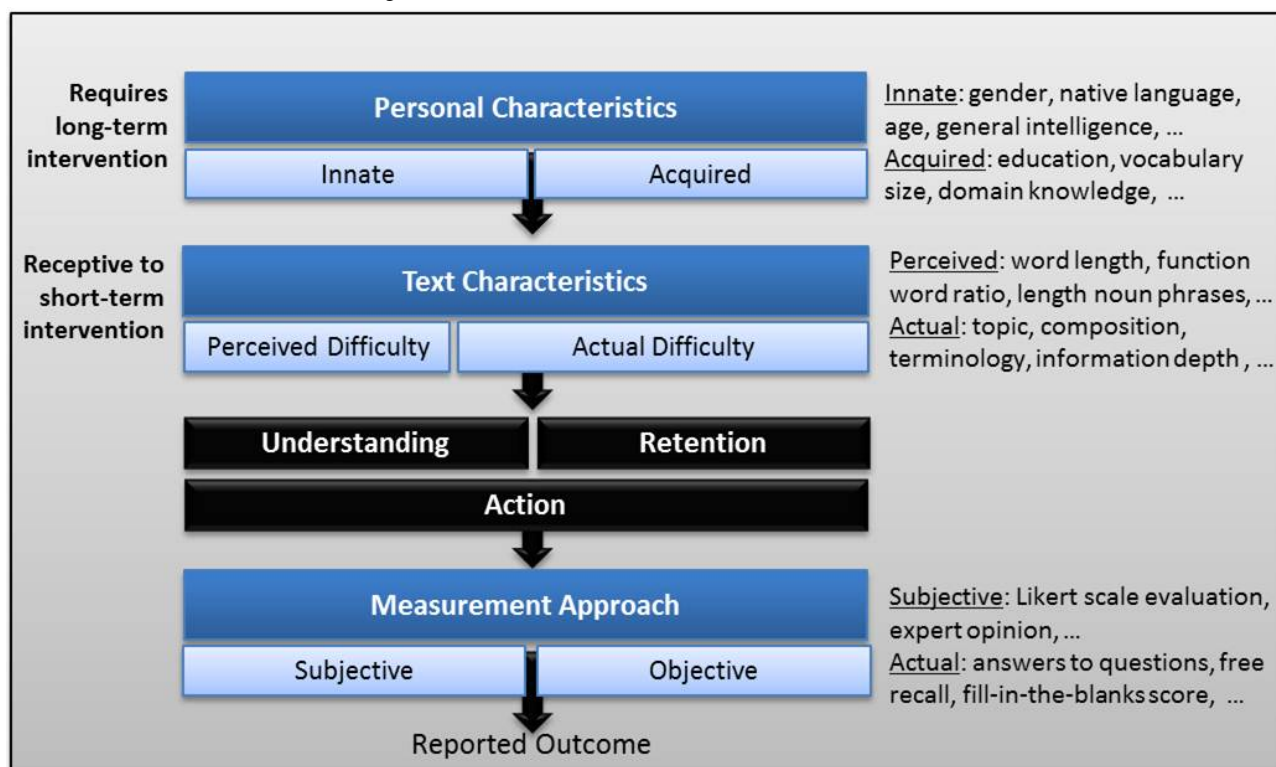
Methods

Text Simplification Algorithm and Writing Process

The automated algorithm executes two steps. The first step is *identification of difficult terms.* We conducted corpus analyses and found that the term familiarity differed between easy and difficult texts [17,18]. Motivated by this, our algorithm uses the Google Web Corpus [37], which contains *n*-gram counts from a corpus of 1 trillion words from public webpages to identify difficult terms. Terms with a low frequency in this corpus are assumed to be less familiar and therefore more difficult since a reader would not encounter them often. We used unigrams and the 5000th most frequent word, which has a frequency of 15,377,914, as our threshold for distinguishing less familiar terms. Any term with a lower frequency is considered difficult and is a candidate for replacement.

We used the Google Web Corpus because its terms are representative of everyday readers without special medical knowledge. Other resources may provide additional value but may also introduce inconsistencies. For example, the Google Book Corpus contains many medical books resulting in higher frequencies for medical terms. The Unified Medical Language System (UMLS) contains both medical and general terms. Distinguishing between them algorithmically would be necessary, which is not an easy task, and may not improve upon the frequency-based approach by much.

The second step is the *identification and presentation of easier alternatives for each difficult term.* The list of candidate replacements is generated from synonyms and hypernyms from WordNet 2.0 [38,39]; definitions and semantic types from the UMLS; and definitions from both the English and Simple English Wiktionaries. Only alternatives that possess the same part of speech based on an automatic tagger are presented. In addition, only substitutions with a higher term frequency than the original word are suggested (ie, more familiar). The number of alternatives provided can be adjusted based on user preference or application; currently, we aim to provide a minimum of 7 alternatives. Candidate replacements are sorted both by source (for the convenience of the writer) and by their familiarity in the Google Web Corpus.

In contrast to the previous version of our simplification algorithm [36], which involved one of the authors manually looking up each word to generate the candidate suggestions, the current version is fully automated. To ensure that the algorithm is sufficiently efficient for later inclusion in a comprehensive tool, we tested its efficiency on Wikipedia articles. We selected 100 conditions randomly (see Multimedia Appendix 1) from a list of diseases provided by the Mayo Clinic. For each disease, we retrieved the corresponding Wikipedia article. The articles were on average 2573 words long. On average, 617 words were tagged as difficult per article, for which easier alternatives were produced by the algorithm where available. The average run time was 37 seconds per document.

Given the difficulty of completely automated translation, especially in domains such as health where information may not be omitted, we require a writer to finalize the text. At present, a Microsoft Excel spreadsheet is generated containing each original sentence from a text, the same sentence with blanks for all difficult words, and alternatives for each difficult word. The alternatives are presented in a column and ordered according to source and term familiarity. The writer chooses the best alternative, replacing it in the original text. Ensuring grammatical correctness (eg, consistent pluralization) is currently the responsibility of the writer.

Original and Simplified Texts (Study Stimuli)

A subject expert (SE), a medical librarian, simplified the texts. To optimize external validity, we worked with one expert to rewrite the text since this is how the final tool will be used. To increase internal validity, we provided the SE with rules to ensure that we measured only the effects resulting from interaction with our algorithm. She was asked to "Try to replace as many words as possible" and when making a replacement "single words can just be replaced but longer fragments should be added before or after the sentence (with some adjustment for flow of text*)"*. The SE served two main roles: (1) to determine if a difficult word flagged by the algorithm needs to be replaced, and (2) for those words requiring replacement, to select an appropriate substitution from the alternatives suggested by the algorithm. If the SE deemed that an appropriate synonym existed for a difficult word in the algorithmically generated options, then the difficult word was simply replaced by the synonym. If the simplification option selected by the SE was not a synonym, it needed to be added to the text so that no original information was deleted from the text. Simplifications containing longer phrases or sentences (eg, from definitions) were added by using parentheses or by adding a separate sentence before or after the target sentence. The text was adjusted by the SE as necessary to create grammatically correct sentences.

In previous work [36], we noticed that lexical simplifications by the authors reduced the flow of the text thereby increasing text difficulty. Therefore, the SE was asked to pay close attention to how alternatives were inserted and to choose the option that resulted in the best flow. If the SE preferred a term other than those suggested by the algorithm, she could add it to the text for familiarity verification. Once the text was rewritten, it was rerun through the simplification algorithm to ensure that newly added text was sufficiently simple. This included the verification of any synonyms by the SE.

To measure *perceived difficulty*, we selected 5 text snippets; these were individual sentences and in one case 2 short sentences combined. Such short snippets do not require much time to read, provide more data points than one long text, and ensure that study participants do not get overwhelmed. The sentences were taken from English Wikipedia articles, and each sentence was simplified by the SE using our algorithm. Our algorithm tagged an average of 11 words per sentence as difficult, of which 5.6 (53%) were replaced.

To measure *actual difficulty,* it was necessary to use longer texts to allow for questions about the content to be posed. We used two different texts so that each participant in the study worked with an original and simplified text for better (statistical) control of interpersonal differences. We chose a text on liver cirrhosis and one on asthma because most people are somewhat familiar

with them and both conditions have several commonly accepted myths associated with them. These myths were incorporated into our multiple-choice questions and provided an excellent opportunity to demonstrate learning. Each text was simplified using our approach described above. Texts were obtained from the initial summary paragraphs from their Wikipedia Web pages and were similar in composition. Our algorithm tagged 210 words as difficult in the liver cirrhosis document, of which 66 (31%) were replaced by the writer during simplification. In the asthma document, 122 words were tagged as difficult and 53 (43%) were replaced during simplification.

Tables 1 and 2 show an overview of the text characteristics before and after simplification. We include the Flesch-Kincaid Grade Level for comparison with other work. Below are examples of an original and simplified snippet used as part of the study (perceived difficulty):

- original: "Gout is a disorder of purine metabolism, and occurs when its final metabolite, uric acid, crystallizes in the form of monosodium urate, precipitating in joints, on tendons, and in the surrounding tissues."
- simplified: "Gout is a disease of the processing of the chemical substance called purine, and occurs when its last chemical product (uric acid) makes crystals (monosodium urate), which collect in joints, on tendons, and in the surrounding tissues."

The texts, both original and simplified versions, are provided in Multimedia Appendix 2.

Table 1. Text snippet characteristics.

	Lexical simplification	
Sentences (N=5)	Original	Simplified
Word count (avg)	28.4	37.6
Flesch-Kincaid grade level (avg)	18.6	17.3

Table 2. Document characteristics.

	Lexical simplification					
Documents	Original			Simplified		
Topic	Asthma	Liver cirrhosis	Average	Asthma	Liver cirrhosis	Average
Word count	623	481	552	779	696	737.5
Sentence count	31	25	28	33	27	30
Flesch-Kincaid grade level	13.9	14.5	14.2	13.7	14.3	14.0

Metrics

To measure *perceived difficulty*, participants judged a sentence using a 5-point Likert scale with the following labels: Very Easy, Easy, Neither, Hard, Very Hard. Perceived difficulty is the score on this scale with 1 representing Very Easy and 5 Very Hard.

To measure *actual difficulty*, we used metrics covering understanding, learning, and retention of information. For understanding of the text, we used 5 multiple-choice questions posed alongside the text. The questions targeted different sections of the text. Understanding was measured as the percentage of questions answered correctly.

To measure learning, we compared scores on 7 multiple-choice questions shown both before and after reading the text. The text itself was not visible when the questions were presented. By asking the same questions before and after, we were able to use participants as their own controls. For each text, we created the multiple-choice questions based on commonly accepted myths. The myths were gathered by searching the Internet for "common myths about…". Learning was measured as the increase in the percentage of questions answered correctly after versus before reading the text.

To measure retention, we asked participants after all sections have been completed to list all facts (one per line) that they remembered from the texts. Retention can be simply measured as the number of facts listed, however, since these facts may contain errors, they were also graded by the authors. Two authors per topic independently graded all facts. Even though participants were asked to list 1 fact per line, many lines included multiple facts per line. Each fact was considered and awarded points separately: +1 for a correct fact and -1 for an incorrect. To grade the answers in an objective manner, the order of answers was randomized per grader and the experimental condition unknown. In cases with a large disparity between grades (scores diverged by more than 100%), a third grader (the SE) judged the results and provided the final score (similar to original manual GRE scoring [40]). Retention was then measured with 2 metrics: the number of listed facts and the sum of the grades assigned to those facts.

In addition to study questions, we also included qualifying questions. These were simple questions for which the answer was obvious. They helped filter results of participants who were not serious about the study. We included a qualifying question with each set of multiple-choice questions and filtered any participant who did not get all qualifying questions right.

Participants

Participants were recruited using Amazon's Mechanical Turk. MTurk is an online crowdsourcing service that allows for small tasks to be accomplished by human workers. Currently, Amazon has over 300,000 requested tasks and over half a million workers. Workers are paid a small sum for each task accomplished. MTurk has been used in a wide range of settings ranging from user studies to data annotation to subjective rating generation [41]. The workers are a diverse group from all over the world with varied demographic characteristics [42,43]. When precautions are taken to filter out ineffective workers, the quality of the data obtained has been shown to be at least as good as data obtained from more traditional approaches [43,44].

Procedure

Participants were directed to our study website from MTurk, and the sections were presented in the following order:

- The *first page* contained the welcome note and instructions to complete the study sections in order and without use of external sources. From this point, the browser back button was disabled.
- The *first study section* showed the myth-based questions for a topic. Then, the text was shown together with new questions, followed by a repetition of the myth-based questions without the text. For each participant, the order of the questions and answers for each question were randomized. The topic was either liver cirrhosis or asthma, and the version was either original or simplified.
- The *second study section* was identical to the first, but with a different text in a different version. Each participant received one original and one simplified version. The order and topics were balanced over the study so that all combinations of topic and difficulty level were presented.
- The *third study section* contained the individual sentences that participants judged for perceived difficulty. The original and simplified version of a sentence were paired because showing all sentences in one list made it very difficult for participants to notice differences and provide a rational judgment. The order within each pair and the order of the 5 pairs were randomized per participant.
- The *fourth study section* contained demographic questions.
- The *fifth and sixth study sections* contained the PSS-10 [45], a standardized stress survey, and the STOFHLA [46,47], a standardized health literacy measure.
- The *seventh and eighth sections* contained the request for free recall of information for the first and second text.
- The *final page* showed a Thank You note and the code to be submitted for payment at MTurk.

Results

Participant Characteristics

We invited MTurk workers located in the United States with a 95% approval rate on tasks previously performed for other requesters. They were paid US$1.50 for completing the survey. Upon start, 134 participants signed up and 105 completed the study. Of those who completed, 6 did not pass our filtering criteria resulting in a total of 99 valid participants. Completing the survey took on average 33 minutes. The shortest time spent was 13 minutes and the longest was 45 minutes.

Table 3 provides the demographic information. Most participants (80%) were between 21 and 50 years old, with only a small group younger than 20 (3%) or older than 60 years (4%). The majority were female (63%), white (89%), and not Hispanic or Latino (93%). Most had moderate education: 48% had a high school diploma, 16% an associate's degree, and 25% a bachelor's degree. The majority (89%) spoke exclusively English at home.

Perceived Difficulty

We found a significant beneficial effect of simplification on perceived difficulty with simplified sentences being judged as simpler. Figure 2 shows an overview of the average score and standard error bars for each sentence and for all sentences combined. A paired-samples *t*-test showed the difference to be significant for all pairs ($P<.001$) and for all pairs combined ($P<.001$).

Table 3. Participant demographic information (n=99).

Characteristics		n
Age		
	20 or younger	3
	21-30	35
	31-40	24
	41-50	21
	51-60	12
	61-70	4
	71 or older	-
Gender		
	Female	62
	Male	37
Race (multiple choices allowed)		
	American Indian / Native Alaskan	2
	Asian	7
	Black or African American	5
	Native Hawaiian or Other Pacific Islander	-
	White	88
Ethnicity		
	Hispanic or Latino	7
	Not Hispanic or Latino	92
Education (highest completed)		
	Less than High School	1
	High School Diploma	48
	Associate's Degree	16
	Bachelor's Degree	25
	Master's Degree	6
	Doctorate	3
Language skills (frequency of speaking English at home)		
	Never English	-
	Rarely English	1
	Half English	3
	Mostly English	6
	Only English	89

Figure 2. Average perceived difficulty scores (lower score = perceived simpler).
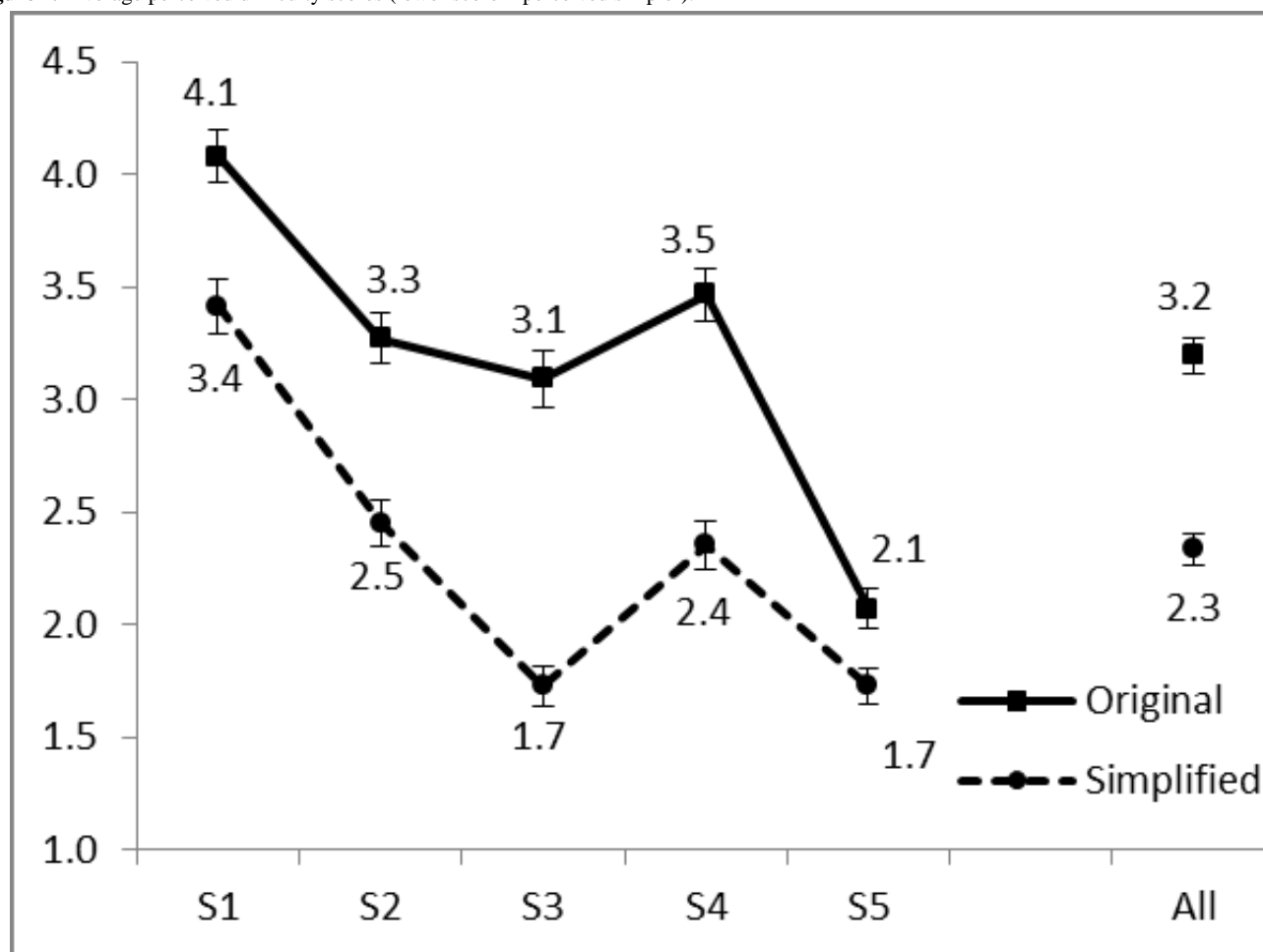
Actual Difficulty: Understanding, Learning, and Retention

Figure 3 shows the mean scores and standard error bars for understanding. We conducted a two-way analysis of variance (ANOVA) with simplification and topic as independent variables. Topic was included to provide a more nuanced view. For understanding, we found two main effects. The first is for simplification with higher scores for simplified text. There were on average 52% correct answers with an original document and 63% with a simplified document, $(F_{1,198}=13.869, P<.001)$. There was also a main effect for topic $(F_{1,198}=13.869, P<.001)$ with higher scores achieved for the asthma document. Since the increases in understanding after simplification were comparable for both topics, the interaction effect was not significant.

Figure 4 shows the mean scores and standard error bars for the learning of information. We conducted a comparable two-way ANOVA with the simplification and topic as independent variables. We found a significant main effect of simplification of text with more learning from simplified documents (18%)

than from the original documents (9%) $(F_{1,198}=9.238, P=.003)$. A second main effect was found for topic $(F_{1,198}=22.301, P<.001)$ with more learning with the liver cirrhosis document (20%) than with the asthma document (6%). The interaction between both independent variables was also significant $(F_{1,198}=4.071, P=.045)$ with the learning being more pronounced with the liver cirrhosis than with the asthma document.

Table 4 provides an overview of the retention of information using both raw and graded scores. With simplified documents, slightly more facts were listed (5.04) than with original documents (4.66). There were also slightly more words (43.60) and unique words (32.36) used after reading simplified documents compared to original documents (40.07 words and 30.79 unique words). These differences were not statistically significant. The graded scores show similar small differences. There were slightly more correct facts after reading simplified documents (5.04 facts) than after reading the original documents (4.70 facts). However, the difference is not statistically significant.

Figure 3. Average understanding scores.
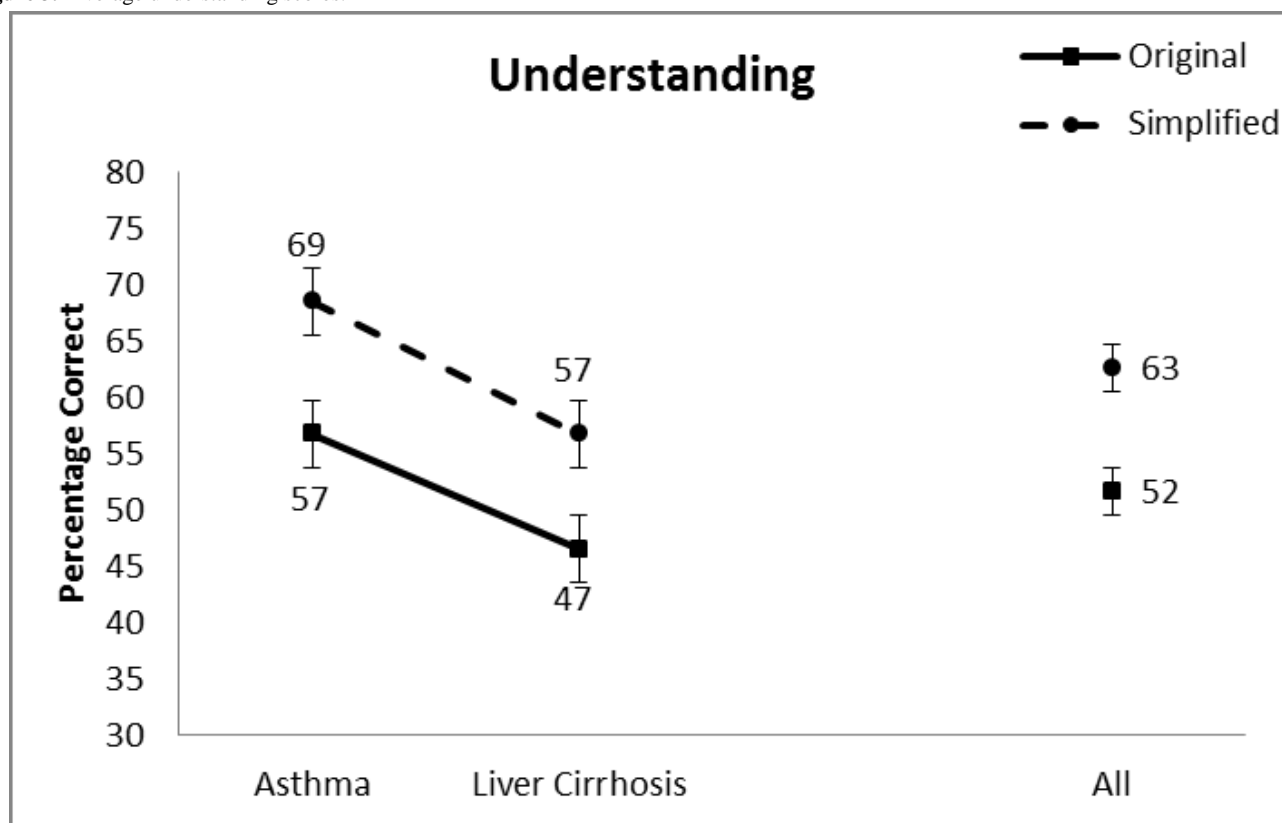
Figure 4. Average learning scores.
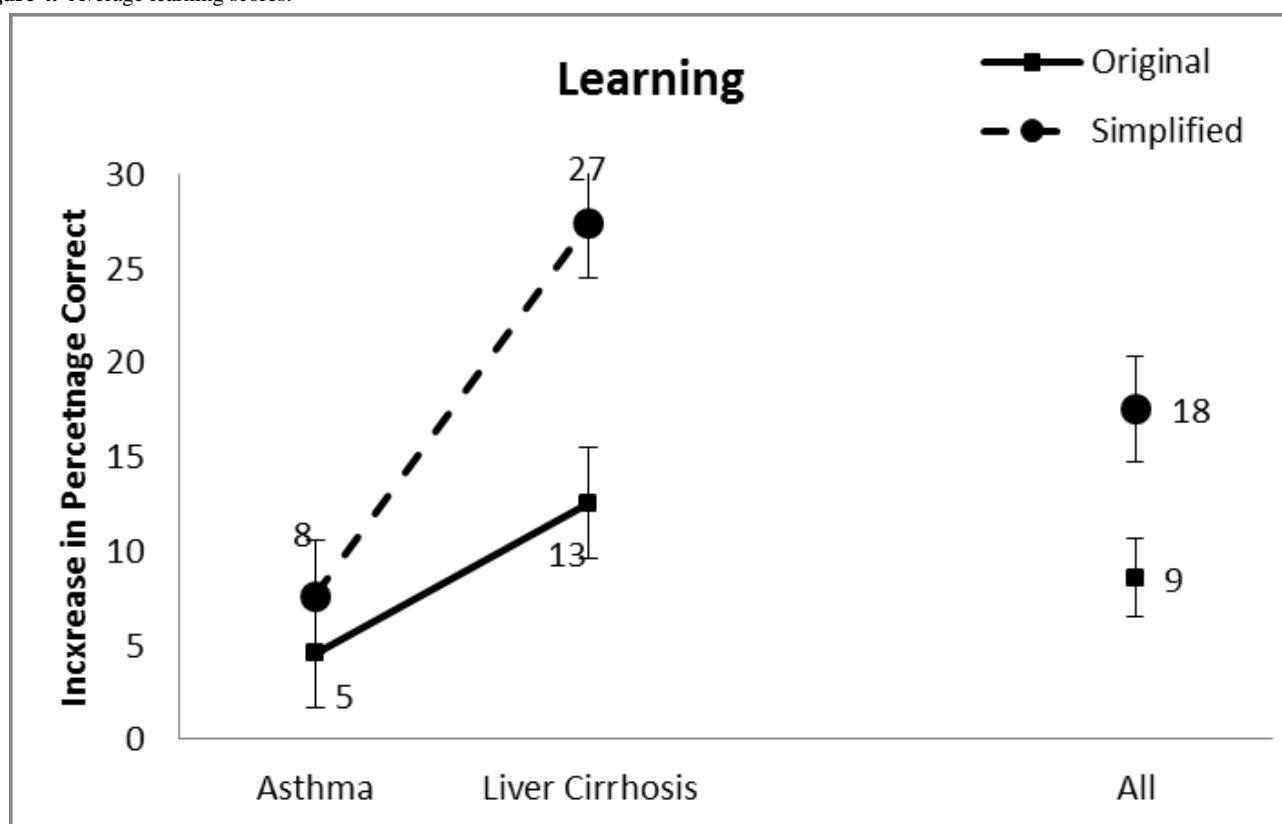
Table 4. Retention of information: mean and standard deviation.

Average counts	Original text			Simplified text			
	Asthma	Liver cirrhosis	Average	Asthma	Liver cirrhosis	Average	
Facts	4.76	4.55	4.66	4.96	5.12	5.04	
Words	40.50	39.63	40.07	42.90	44.24	43.60	
Unique words	31.38	30.20	30.79	32.39	32.34	32.36	
Average score (graded facts)	4.50	4.91	4.70	4.73	5.35	5.04	

Relationships With Participant Characteristics

To complete our analysis, we conducted a correlation analysis using a 2-tailed Pearson product-moment correlation coefficient (*r*). We evaluated the personal characteristics and the scores for perceived and actual difficulty over experimental conditions. We assigned a code to the education level and language skills with a higher score indicating a higher level or skill. We also included the PSS scores and STOFHLA scores.

Overall, there were few significant correlations. There were no significant correlations between the perceived difficulty of sentences and the personal characteristics. For actual difficulty, only education mattered. There was a positive correlation between education and understanding (r=.244, P=.015), facts listed (r=.296, P=.003), graded facts (r=.411, P≤.001), and both the word count (r=.316, P=.001) and unique word count (r=.329, P=.001). Among the personal characteristics themselves, two correlations were significant. There was a negative correlation between language skills and stress levels, indicating higher stress related to lower language skills (r=-.210, P=.037) and also a negative correlation between language skills and education level (r=-.260, P=.009). Upon closer inspection, this last negative correlation was due to a few individuals with higher degrees who speak a different language at home, that is, Chinese, Tamil, or Farsi.

Discussion

Principal Findings

This work reported on a lexical simplification algorithm that automatically detects difficult terms and suggests easier alternatives. The writing process is semiautomated since the final replacements are made by the writer. A controlled user study showed how simplifying text in this manner led to significant improvements in both perceived and actual difficulty of text.

The results on perceived difficulty corroborate earlier work on manual lexical simplification. In general, changing the text to improve perceived difficulty is more straightforward. Consistent and strong effects are found even when using short text snippets or small sample sizes. Even so, this effect is important and shows that *lexical simplification has a beneficial impact on perceived difficulty*. Future studies will look more closely at how perceived difficulty affects motivation to read and ability to complete reading, among other factors.

The results on actual difficulty are strong and very encouraging. They also show the importance of using different metrics. We

found a strong effect on understanding with simplified text being better understood. However, this effect also depended on the topic being studied. Learning showed a similar strong effect: there was more learning with simplified documents. These effects lead to our conclusion that *lexical simplification is beneficial and has an immediate impact on understanding and learning*. However, we did not find an effect of simplification on retention of information. This may be due to a lack of sustained learning or it may be due to the study design. In previous work on search engines [48], we found that many study participants stop finding information at some given point, regardless of how easy or difficult a task is. We may be witnessing a similar effect with participants submitting "enough" facts regardless of how many they remember. In future work, we aim to provide better incentives to encourage participants to submit more facts.

Limitations

There are several limitations we would like to point out. First, we evaluated our approach with short texts taken from Wikipedia. Different effects may be found for longer or more difficult texts. However, working with short texts allows for a controlled experiment, thereby avoiding potentially confounding variables. Future work will look for repeat effects in longer documents. Second, we worked with general topics. Automatically recognizing which different texts, either distinguished by difficulty level or other factors, would benefit from simplification would be an important addition to our work. In addition, working with personally relevant topics may increase effects, since motivation has been shown to be important to the reading process. Third, we worked with only one subject expert who rewrote text. Comparing different writers may show further strengths and weaknesses of our approach. Working with a team of writers may provide a more balanced gold standard; however, this approach has also been shown to introduce noise when experts disagree [49]. Further research is needed to understand the impact of each of these limitations.

Conclusions

In addition to these study limitations, there is also much room for future development of our algorithm. We aim to more precisely target difficult words so that fewer words are tagged for replacement while still impacting the overall difficulty of text. We aim to provide a shorter and more precise list of potential replacements by working with resources such as the Consumer Health Vocabulary [50-52]. This will make the process more efficient for the writer while requiring less time to generate alternatives. For example, we plan to test phrases in addition to individual words to estimate difficulty and work

with different thresholds. We also are working toward combining lexical simplification with other forms of simplification of relevant text features.

Acknowledgments

Conflicts of Interest

None declared.

Multimedia Appendix 1

List of conditions.

[PDF File (Adobe PDF File), 7KB - jmir_v15i7e144_app1.pdf]

Multimedia Appendix 2

Stimuli: Original and simplified texts.

[PDF File (Adobe PDF File), 62KB - jmir_v15i7e144_app2.pdf]

References

1. Fox S. Health Topics.: Pew Research Center; 2011. URL: http://www.pewinternet.org/Reports/2011/HealthTopics.aspx [accessed 2013-07-07] [WebCite Cache ID 6HwPg8amf]

2. Wang Y. Automatic Recognition of Text Difficulty from Consumers Health Information. In: Proceedings of the 19th IEEE International Symposium on Computer-Based Medical Systems. 2006 Presented at: 19th IEEE International Symposium on Computer-Based Medical Systems; June 22-23, 2006; Salt Lake City.

3. Nielsen-Bohlman L, Institute of Medicine /Committee on Health. 'Health literacy: a prescription to end confusion'. Maryland: National Academies Press; 2004.

4. Yan X, Song D, Li X. Concept-based Document Readability in Domain Specific Information Retrieval. In: Proceedings of the 15th ACM International Conference on Information and Knowledge Management. 2006 Presented at: 15th ACM International Conference on Information and Knowledge Management; Nov. 5-11, 2006; Arlington, Virginia p. 540-549.

5. Zarcadoolas C. The simplicity complex: exploring simplified health messages in a complex world. Health Promot Int 2011 Sep;26(3):338-350 [FREE Full text] [doi: 10.1093/heapro/daq075] [Medline: 21149317]

6. Low health literacy report. URL: http://sphhs.gwu.edu/departments/healthpolicy/CHPR/downloads/LowHealthLiteracyReport10_4_07.pdf [accessed 2013-02-06] [WebCite Cache ID 6EEgqYeox]

7. Diakidoy IA, Mouskounti T, Ioannides C. Comprehension and Learning from Refutation and Expository Texts. Reading Research Quarterly 2011;43(1):22. [doi: 10.1598/RRQ.46.1.2]

8. Leroy G, Miller T. Perils of providing visual health information overviews for consumers with low health literacy or high stress. J Am Med Inform Assoc 2010;17(2):220-223 [FREE Full text] [doi: 10.1136/jamia.2009.002717] [Medline: 20190068]

9. Van Servellen G, Brown JS, Lombardi E, Herrera G. Health literacy in low-income Latino men and women receiving antiretroviral therapy in community-based treatment centers. AIDS Patient Care and STDs 2003 Jun;17(6):283-298. [doi: 10.1089/108729103322108166] [Medline: 12880492]

10. Kintsch W. Learning from Text. Cognition and Instruction 1986;3(2):87.

11. Landi N. An examination of the relationship between reading comprehension, higher-level and lower-level reading sub-skills in adults. Read Writ 2010 Jul 1;23(6):701-717 [FREE Full text] [doi: 10.1007/s11145-009-9180-z] [Medline: 21691452]

12. Stanovich KE, West RF. Exposure to Print and Orthographic Processing. Reading Research Quarterly 1989;24(4):402.

13. Leroy G, Lauchak D, Mouradi O. A User-study Measuring the Effects of Lexical Simplification and Coherence Enhancement on Perceived and Actual Text Difficulty. Int J Med Inform 2013 (forthcoming).

14. Janz NK, Becker MH. The Health Belief Model: a decade later. Health Educ Q 1984;11(1):1-47. [Medline: 6392204]

15. Trafimow D, Sheeran P, Conner M, Finlay KA. Evidence that perceived behavioural control is a multidimensional construct: perceived control and perceived difficulty. Br J Soc Psychol 2002 Mar;41(Pt 1):101-121. [Medline: 11970777]

16. Velayo RS. Retention of Content as a Function of Presentation Mode and Perceived Difficulty. Reading Improvement 1993;30(4):216.

17. Leroy G, Endicott J. Term Familiarity to Indicate Perceived and Actual Difficulty of Text in Medical Digital Libraries. In: Proceedings of the International Conference on Asia-Pacific Digital Libraries (ICADL) - Digital Libraries -- for Culture

Heritage, Knowledge Dissemination, and Future Creation. 2011 Presented at: International Conference on Asia-Pacific Digital Libraries (ICADL) - Digital Libraries -- for Culture Heritage, Knowledge Dissemination, and Future Creation; October 2011; Beijing, China p. 24-27.

18. Leroy G, Endicott J. Combining NLP with Evidence-based Methods to Find Text Metrics related to Perceived and Actual Text Difficulty. In: Proceedings of the 2nd ACM SIGHIT International Health Informatics Symposium (ACM IHI). 2012 Presented at: 2nd ACM SIGHIT International Health Informatics Symposium (ACM IHI); January 2012; Miami, FL p. 28-30.

19. Pitler E, Nenkova A. Revisiting Readability: A Unified Framework for Predicting Text Quality. In: Proceedings of the Empirical Methods in Natural Language Processing. 2008 Presented at: Empirical Methods in Natural Language Processing; 2008; Honolulu.

20. Soleimani H, Mohammadi E. The Effect of Text Typographical Features on Legibility,Comprehension, and Retrieval of EFL Learners. English Language Teaching 2012;5(8):207.

21. Goldman SR, Murray JD. Knowledge of Connectors as Cohesion Devices in Text: A Comparative Study of Native-English and English-as-a-Second-Language Speakers. Journal of Educational Psychology 1992;84(4):504.

22. Boscolo P, Mason L. Topic Knowledge, Text Coherence, and Interest: How they Interact in Learning from Instructional Texts. The Journal of Experimental Education 2003;7(2):126.

23. McNamara DS, Kintsch E, Songer NB, Kintsch W. Are Good Texts Always Better? Interactions of Text Coherence, Background Knowledge, and Levels of Understanding in Learning from Text. Cognition and Instruction 1996;14(1):1.

24. DuBay WH. The Principles of Readability. Impact Information 2004.

25. Bruce B, Rubin A, Starr K. Why Readability Formulas Fail. IEEE Transactions on Professional Communication 1981;24(1):50.

26. Wang LW, Miller MJ, Schmitt MR, Wen FK. Assessing Readability Formula Differences with Written Health Information Materials: Application, Results, and Recommendations. Research in Social & Administrative Pharmacy 2012 (forthcoming).

27. Sirico LJ. Readability Studies: How Technocentrism Can Compromise Research and Legal Determinations. Villanova University Legal Working Paper Series 2008.

28. Polishchuk DL, Hashem J, Sabharwal S. Readability of online patient education materials on adult reconstruction Web sites. J Arthroplasty 2012 May;27(5):716-719. [doi: 10.1016/j.arth.2011.08.020] [Medline: 22000573]

29. Ahmed OH, Sullivan SJ, Schneiders AG, McCrory PR. Concussion information online: evaluation of information quality, content and readability of concussion-related websites. Br J Sports Med 2012 Jul;46(9):675-683. [doi: 10.1136/bjsm.2010.081620] [Medline: 21504964]

30. Bailey M, Coughlin PA, Sohrabi S, Griffin KJ, Rashid ST, Troxler MA, et al. Quality and readability of online patient information for abdominal aortic aneurysms. J Vasc Surg 2012 Jul;56(1):21-26. [doi: 10.1016/j.jvs.2011.12.063] [Medline: 22521801]

31. Hendrick PA, Ahmed OH, Bankier SS, Chan TJ, Crawford SA, Ryder CR, et al. Acute low back pain information online: an evaluation of quality, content accuracy and readability of related websites. Man Ther 2012 Aug;17(4):318-324. [doi: 10.1016/j.math.2012.02.019] [Medline: 22464886]

32. Kim H, Goryachev S, Rosemblat G, Browne A, Keselman A, Zeng-Treitler Q. Beyond Surface Characteristics: A New Health Text-Specific Readability Measurement. AMIA Annu Symp Proc 2007:418-422.

33. Gemoets D, Rosemblat G, Tse T, Logan R. Assessing readability of consumer health information: an exploratory study. Stud Health Technol Inform 2004;107(Pt 2):869-873. [Medline: 15360936]

34. Clinton V, Broek P. Interest, inferences, and learning from texts. Learning and Individual Differences 2012;22(6):650.

35. Erçetin G. Effects of Topic Interest and Prior Knowledge on Text Recall and Annotation use in Reading a Hypermedia Text in the L2. ReCALL 2010;22(2):228.

36. Leroy G, Endicott J, Mouradi O, Kauchak D, Just M. Improving Perceived and Actual Text Difficulty for Health Information Consumers using Semi-Automated Methods. In: Proceedings of the American Medical Informatics Association (AMIA) Fall Symposium. 2012 Presented at: American Medical Informatics Association (AMIA) Fall Symposium; Nov. 3-7, 2012; Chicago, IL.

37. Web 1T 5-gram Corpus Version 1.1. URL: http://www.ldc.upenn.edu/Catalog/docs/LDC2006T13/readme.txt [accessed 2013-07-07] [WebCite Cache ID 6HwSRwun2]

38. Fellbaum C. WordNet: An Electronic Lexical Database. Cambridge, MA: MIT Press; 1998.

39. Miller GA. WordNet: a lexical database for English. Commun. ACM 1995;38(11):39.

40. How the GRE tests are scored. URL: http://www.ets.org/gre/institutions/scores/how [accessed 2013-02-06] [WebCite Cache ID 6EEh5ILJE]

41. Kittur A, Chi E, Suh B. Crowdsourcing User Studies with Mechanical Turk. In: Proceedings of the SIGCHI Conference on Human Factors in System Computing. 2008 Presented at: Proceedings of the SIGCHI Conference on Human Factors in System Computing; 2008; Florence, Italy p. 453-456.

42. Ross J, Irani L, Silberman MS, Zaldivar A, Tomlinson B. Who are the Crowdworkers? Shifting Demographics in Mechanical Turk. In: Proceedings of the CHI '10 Extended Abstracts on Human Factors in Computing Systems. 2010 Presented at: CHI '10 Extended Abstracts on Human Factors in Computing Systems; 2010; Atlanta, Georgia.

XSL·FO
RenderX

43. Paolacci G, Changler J, Ipeirotis PG. Running Experiments on Amazon Mechanical Turk. Judgment and Decision-making 2010;5(5):411.

44. Buhrmester M, Kwang T, Gosling SD. Amazon's Mechanical Turk A New Source of Inexpensive, Yet High-Quality, Data? Perspectives on Psychological Science 2011;6(1).

45. Cohen S, Kamarck T, Mermelstein R. A Global Measure of Perceived Stress. Journal of Health and Social Behavior 1983;24(4):385.

46. Nurss JR, Parker RM, Williams MV, Baker DW. Test of Functional Health Literacy in Adults. Hartford, MI: Peppercorn Books & Press; 1995.

47. Parker RM, Baker DW, Williams MV, Nurss JR. The Test of Functional Health Literacy in Adults: A New Instrument for Measuring Patients' Literacy Skills. Journal of General Internal Medicine 1995;10:537-541.

48. Leroy G, Lally A, Chen H. The Use of Dynamic Contexts to Improve Casual Internet Searching. ACM Transactions on Information Systems 2003;21(3):229.

49. Leroy G, Rindflesch TC. Effects of information and machine learning algorithms on word sense disambiguation with small datasets. Int J Med Inform 2005 Aug;74(7-8):573-585. [doi: 10.1016/j.ijmedinf.2005.03.013] [Medline: 15897005]

50. Zeng-Treitler Q, Goryachev S, Tse T, Keselman A, Boxwala A. Estimating consumer familiarity with health terminology: a context-based approach. J Am Med Inform Assoc 2008;15(3):349-356 [FREE Full text] [doi: 10.1197/jamia.M2592] [Medline: 18308983]

51. Keselman A, Smith CA, Divita G, Kim H, Browne A, Leroy G, et al. Consumer Health Concepts that do not Map to the UMLS: Where Do They Fit? Journal of the American Medical Informatics Association 2008;15(4):496-505.

52. Keselman A, Logan R, Smith CA, Leroy G, Zeng-Treitler Q. Developing informatics tools and strategies for consumer-centered health communication. J Am Med Inform Assoc 2008;15(4):473-483 [FREE Full text] [doi: 10.1197/jamia.M2744] [Medline: 18436895]

Abbreviations

ANOVA: analysis of variance
GRE: Graduate Record Examinations
SE: subject expert
UMLS: Unified Medical Language System

XSL•FO
RenderX

Original Paper

Understanding Patient Portal Use: Implications for Medication Management

Chandra Y Osborn[1,2,3], PhD, MPH; Lindsay Satterwhite Mayberry[1], PhD, MS; Kenneth A Wallston[4], PhD; Kevin B Johnson[2,5], MD, MS; Tom A Elasy[1,3], MD, MPH

[1]Department of Medicine, Vanderbilt University Medical Center, Nashville, TN, United States

[2]Department of Biomedical Informatics, Vanderbilt University Medical Center, Nashville, TN, United States

[3]Diabetes Research & Training Center, Center for Diabetes Translational Research, Vanderbilt University Medical Center, Nashville, TN, United States

[4]School of Nursing, Vanderbilt University Medical Center, Nashville, TN, United States

[5]Department of Pediatrics, Vanderbilt University Medical Center, Nashville, TN, United States

Corresponding Author:
Chandra Y Osborn, PhD, MPH
Department of Medicine
Vanderbilt University Medical Center
1215 Twenty-First Ave South
Ste 6000, MCE - North Tower
Nashville, TN, 37232-8300
United States
Phone: 1 615 936 8468
Fax: 1 615 936 1269
Email: chandra.osborn@vanderbilt.edu

Abstract

Background: The Internet can be leveraged to provide disease management support, including medication adherence promotion that, when tailored, can effectively improve adherence to medications. The growing adoption of patient portals represents an opportunity to support medication management and adherence more broadly, but virtually no data exist about the real and potential impact of existing portals on these outcomes.

Objective: We sought to (1) understand who uses an existing patient portal and reasons for use and nonuse, (2) understand how portal users are using a portal to manage their medications, and (3) explore participants' ideas for improving portal functionality for medication management and adherence support.

Methods: A total of 75 adults with type 2 diabetes participated in a mixed-methods study involving focus groups, a survey, and a medical chart review. We used quantitative data to identify differences between portal users and nonusers, and to test the relationship between the frequency of portal use and glycemic control among users. We used qualitative methods to understand how and why participants use a portal and their ideas for improving its medication management functionality.

Results: Of the enrolled participants, 81% (61/75) attended a focus group and/or completed a survey; portal users were more likely than nonusers to participate in that capacity (Fisher exact test; P=.01). Users were also more likely than nonusers to be Caucasian/white (Fisher exact test; P<.001), have higher incomes (Fisher exact test; P=.005), and be privately insured (Fisher exact test; P<.001). Users also tended to have more education than nonusers (Mann–Whitney U; P=.05), although this relationship was not significant at P<.05. Among users, more frequent use of a portal was associated with better A1C (Spearman rho =–0.30; P=.02). Reasons for nonuse included not knowing about the portal (n=3), not having access to a computer (n=3), or having a family member serve as an online delegate (n=1). Users reported using the portal to request prescription refills/reauthorizations and to view their medication list, and they were enthusiastic about the idea of added refill reminder functionality. They were also interested in added functionality that could streamline the refill/reauthorization process, alert providers to fill/refill nonadherence, and provide information about medication side effects and interactions.

Conclusions: Although there are disparities in patient portal use, patients use portals to manage their medications, are enthusiastic about further leveraging portals to support medication management and adherence, and those who use a portal more frequently have better glycemic control. However, more features and functionality within a portal platform is needed to maximize medication management and adherence promotion.

XSL•FO
RenderX

Introduction

Diabetes pharmacotherapy improves glycemic control and prevents diabetes-related complications [1], but many individuals with type 2 diabetes mellitus (T2DM) are nonadherent to prescribed medications [2-4]. Medication nonadherence includes not taking the appropriate dose of a medication or not taking it at the correct time (ie, suboptimal dose adherence), abandoning a medication all together, or not picking up or reauthorizing an existing medication (ie, suboptimal refill adherence) [5]. Optimal diabetes self-care is multifaceted, requiring physical activity, appropriate nutrition, blood glucose monitoring, and medication adherence [6], and yet suboptimal medication adherence alone is independently associated with poor glycemic control [4,7], an increased risk of hospitalizations and mortality [8,9], and higher health care costs [9].

The Internet can provide a platform for disease management support [10], including medication adherence promotion [11,12] that, when tailored, effectively improves adherence to long-term medications [11]. Patient portals are secure Internet-based platforms that offer patients the ability to view their personal health information (PHI), and some portals also allow for 2-way secure messaging between patients and health care providers, and the ability to schedule medical appointments and request prescription refills [13-15]. Emerging evidence suggests portals improve health care quality [16,17] and are associated with favorable patient outcomes [18,19]. For example, among patients with diabetes, using a portal or a comparable system has been associated with performing self-care activities and glycemic control [18,19]. On the basis of this and what we know about the efficacy of Internet-based interventions to promote medication adherence [11], offering medication management support via a patient portal [20] may be an effective means of promoting medication adherence to a large audience over a sustained period of time.

In recent years, patient portal functionality has become more robust. Portals are not only allowing patients to perform the tasks described previously, but they also allow patients to transfer, share, and print PHI [21]; receive generic health information [22] and/or personally relevant health information [22,23]; self-screen for acute health issues (eg, flu) [24]; be coached on how to communicate with providers before medical appointments [25]; and manage medication lists [21,26] among other medication management tasks. From 1 of these studies, we learned that having access to personally relevant health information promotes sustained use of a patient portal [22]. In addition, patients who had access to a medication management module added to an existing patient portal were both willing to use and satisfied with using it to reconcile medication lists and to report side effects, adverse drug events, and issues pertaining to medication nonadherence [20]. However, we know very little about how patients are using standard patient portal functionality for medication management and adherence support purposes and what types of tools are currently absent from these platforms that could be added to meet the medication adherence-related needs of patients.

In an effort to learn how individuals with a chronic illness use patient portal technology to manage their medications, we conducted a mixed-methods study with adults with a T2DM diagnosis who had been prescribed glucose-lowering medications and/or insulin. Our objectives were to (1) understand who uses an existing patient portal and reasons for use and nonuse, (2) understand how portal users are using a portal to manage their medications, and (3) explore participants' ideas for improving portal functionality for medication management and adherence support.

Methods

MyHealthAtVanderbilt

MyHealthAtVanderbilt (MHAV) is a patient portal available to patients receiving care at Vanderbilt University Medical Center (VUMC), a large academic medical center in Nashville, Tennessee. Users of MHAV can manage medical bills, view PHI (eg, vital signs, laboratory results, medication lists, and diagnoses) from their electronic health record (EHR), use secure messaging to communicate with providers and manage medical appointments, access VUMC information (eg, maps, provider information, and telephone directory), use health screening tools to assess symptoms or risk for various conditions [24], view opportunities to participate in research studies, and view educational content linked to the *International Classification of Diseases, Ninth Revision* (*ICD-9*) codes from their EHR [14,23]. There are 2 levels of MHAV access to protect the confidentiality of PHI: upon registering online, patients receive access to certain MHAV functions, but they cannot view PHI until they provide in-person identification at a VUMC clinic [14].

Participants and Recruitment

To both learn about patients' experiences with using MHAV to manage their health and medication regimens and get their ideas for leveraging this portal to improve medication management and adherence, we recruited English-speaking adults prescribed medications for T2DM who were patients at VUMC primary care clinics. Recruitment included approaching patients in the clinic waiting room, posting flyers advertising the study, and announcing the study on a VUMC listserv. Interested and eligible patients were identified, and completed informed consent procedures to enroll in the study. All enrolled participants consented to having study personnel review their EHR to collect demographic and clinical information (including medication lists), responded to a question about MHAV use, and were invited to attend a focus group and complete a survey. Enrolled participants who did not attend a focus group were invited to an average of 2.9 (SD 0.5) focus group sessions

scheduled at different times of the day. Participants who could not attend a focus group were asked to complete a survey by phone or mail. The Institutional Review Board at Vanderbilt University approved all procedures prior to study enrollment.

Qualitative Data

We conducted 11 focus groups with 2 to 6 participants each, a trained facilitator (authors CYO or LSM), and a trained note taker who recorded verbal and nonverbal communications. All focus group sessions were recorded and transcribed verbatim, using session notes to link participants' comments to survey data. Focus group questions pertained to patients' experiences with diabetes medications, experiences with and attitudes toward using MHAV and other technologies to manage diabetes and medication regimens, and ideas for leveraging MHAV and other technologies to improve medication management and adherence. The primary questions of interest were:

1. Do you use MHAV to manage your diabetes and medication regimens? How/why not?
2. What content, resources, or tools/functions could be added to MHAV to help you manage your medications?
3. What do you think about using MHAV to receive an email reminder when it's time to refill a prescription or receive text messages when it's time to take a dose of medication or insulin (ie, dose reminders)? Would these functions be helpful to you?

We stratified sessions by self-reported frequency of patient portal use at enrollment (described subsequently) to homogenize groups relative to the subject matter and elicit different perspectives [27], facilitate a more comfortable discussion about experiences with technology, and understand similarities and differences across types of users rather than to generate thematic saturation within each user group. We conducted 2 focus groups with nonusers, 5 with low users, and 4 with high users. In focus group sessions with nonusers, we asked if participants were aware of the MHAV patient portal. We then showed a video demonstration of MHAV [28], and asked participants if they thought they would want to use MHAV in the future. If not, they were asked why not; if so, they were asked which of the features showcased in the demonstration video would they use and what other features would they like to see added.

Quantitative Data

Demographics

We reviewed the EHR to collect each enrolled participant's age, gender, and race. Those who responded to a survey also supplied their education, income, and insurance status.

Patient Portal Use

At enrollment, we asked participants how often they use MHAV to manage their health on a scale from 1=not at all to 5=very often. We used the response to this question to categorize enrolled participants as portal users or nonusers, and to operationalize users' frequency of use. For these analyses, those who answered "not at all" were considered nonusers and all others were considered users. We also asked users how long they had used the portal in months and years.

Clinical Characteristics

For all enrolled participants, we reviewed the EHR to collect the number and type of prescribed medications and the most recent glycosylated hemoglobin (A1C) value to assess glycemic control. Those who responded to a survey also supplied their duration of diabetes in months and years as well as their height and weight used to calculate body mass index.

Analyses

All statistical tests were performed using Stata 12 (StataCorp LP, College Station, TX, USA). Descriptive statistics characterized the sample. Mann–Whitney U tests and Fisher exact tests examined group differences between those who participated in a focus group and/or completed a survey (n=61) and those who were enrolled only (n=14) on demographics, MHAV use, and A1C, and then between patient portal users (n=62) and nonusers (n=13) on all variables. Next, Mann–Whitney U tests and Spearman rho (ρ) correlation coefficients tested the relationships between the frequency of MHAV use and demographics and A1C among MHAV users.

We used NVivo 9 (QSR International, Burlington, MA, USA) to code focus group transcripts. The purpose of our analytic approach was not to reach thematic saturation, but rather to explore participants' receptiveness to using a patient portal for medication management and adherence, and to generate ideas for how to tailor technologies to meet the needs of patients with T2DM. First, author LSM read transcripts in their entirety, identifying statements pertaining to participants' opinions about, experiences with, and ideas for using MHAV for medication management and adherence. Units of analysis consisted of statements by single participants and, largely, multiparticipant conversations during which participants built on each other's ideas, interrupted, offered suggestions, and/or indicated a similar or different experience. Next, authors CYO and LSM iteratively reviewed, integrated, and discussed these data until subthemes emerged.

Quality Assurance

We took several steps to enrich the quality of our data and ensure the trustworthiness of our coding process. We participated in, recorded, and analyzed debriefing sessions after each focus group [29]. We stratified focus group sessions by self-reported patient portal use to obtain and compare different perspectives (ie, triangulation of sources) [30], and we used analyst triangulation to explore different interpretations of these data [30].

Results

We enrolled 75 adults with T2DM with a mean age of 56.9 years (SD 8.8); 67% were female, 63% were Caucasian/white, and 33% were African American/black. See Table 1 for additional summary statistics. Of the enrolled participants, 81% attended a focus group session that included a survey (n=45) or completed a survey by phone/mail (n=16). Nonusers of MHAV were less likely than users of MHAV to participate in a focus group or complete a survey (Fisher exact test; P=.01), but there were no differences between focus group/survey

participants and enrolled-only participants on age, race, gender, or A1C.

Table 1. Participant demographic and clinical characteristics stratified by patient portal use.

Demographics	Patient portal use			P value[a]
	Nonusers (n=13)	Users (n=62)	Full sample (N=75)	
Age (years), mean (SD)	58.8 (10.9)	56.5 (8.4)	56.9 (8.8)	.52
Gender, n (%)				.20
Male	2 (15.4)	23 (37.1)	25 (33.3)	
Female	11 (84.6)	39 (62.9)	50 (66.7)	
Race, n (%)				<.001
Caucasian/white	1 (7.7)	46 (74.2)	47 (62.7)	
African American/black	11 (84.6)	14 (22.6)	25 (33.3)	
Education[b] (years), mean (SD)	13.2 (1.8)	15.2 (2.3)	15.0 (2.4)	.05
Income[b] (US $), n (%)				.005
≤39,999	6 (85.7)	12 (23.5)	18 (31.6)	
40,000-59,999	0 (0.0)	15 (29.4)	15 (26.3)	
≥60,000	1 (14.3)	24 (47.1)	24 (42.1)	
Insurance status[b], n (%)				<.001
Private	1 (14.3)	47 (87.0)	48 (78.7)	
Public	5 (71.4)	6 (11.1)	11 (18.0)	
None	1 (14.3)	1 (1.9)	2 (3.3)	
Number of diabetes medications, mean (SD)	1.3 (0.7)	1.3 (0.8)	1.3 (0.8)	.84
Type of diabetes medications, n (%)				.56
Oral agents only	11 (84.6)	40 (64.5)	51 (69.9)	
Insulin only	1 (7.7)	8 (12.9)	9 (12.3)	
Both	1 (7.7)	12 (3.3)	13 (17.8)	
Diabetes duration[b] (years), mean (SD)	7.8 (7.5)	8.0 (6.0)	8.0 (6.1)	.72
Body mass index[b], mean (SD)	39.2 (12.7)	34.4 (10.2)	35.0 (10.5)	.35
A1C (%), mean (SD)	7.1 (1.6)	7.2 (1.6)	7.3 (1.6)	.71

[a]We conducted Fisher exact tests for categorical variables and Mann–Whitney U tests for continuous variables.

[b]Variable collected by survey (n=61).

Who Uses MyHealthAtVanderbilt?

Of all 75 participants who were enrolled in the study, 83% (n=62) were patient portal users and reported using MHAV sometimes to often (mean 3.6, SD 1.0) [31]. Most portal users (72%) reported at least 1 year of use. As shown in Table 1, MHAV users were more likely than nonusers to be Caucasian/white, have higher incomes, and be privately insured. MHAV users also tended to have more education than nonusers (Mann-Whitney U; P=.05) though this relationship was not significant at $P<$.05. We found no other differences between MHAV users and nonusers. Among MHAV users, frequency of use was unrelated to race or indicators of socioeconomic status (SES; ie, education, income, and insurance status). Although there was no difference in A1C between MHAV users and nonusers, more frequent use of the portal was associated with better A1C (ρ=–0.30, P=.02) among users.

Why Do Some Participants Not Use MyHealthAtVanderbilt?

After we showed nonusers the MHAV demonstration video, 4 of 7 nonusers reported they were interested in using the portal, but either had never heard about it, or had heard about it but did not know what its capabilities were. One of these participants said she would certainly use MHAV if she could use a computer:

> *If I knew how to use a computer, I would use [MHAV]. Because, well, really I think all of it would be helpful. [Scheduling] the doctor's appointments, paying the bills—about the medications—[being able to request] different medications if you need to do*

that, and knowing about your test results. [56-year-old female, African American/black, nonuser]

In general, nonusers who were interested in using MHAV reported wanting to use the secure messaging feature to schedule appointments and wanting to view their PHI, such as their laboratory results and medication list.

Of the 3 participants who were not interested in using MHAV, 2 participants did not know how to use a computer and felt they had good systems in place for managing their health and medications, and the third participant reported her husband used MHAV and managed her medications as her online delegate (see [14,31] for more detail on the delegate function of MHAV).

How Portal Users Use MyHealthAtVanderbilt to Manage Their Medications

Themes and subthemes are illustrated with quotations in the text and in Table 2. In focus groups, users described using MHAV to review their medication list and request prescription reauthorizations. Participants who used MHAV to view their medication list shared the information with other providers and/or pharmacists and used this information to ensure they were taking medications correctly:

When I visited the physician about my injury, I thought she suggested I take 600 mg of Motrin for 2 weeks. However, when I looked at my medication list, it showed 400 mg of Motrin twice a day. It was helpful to see what she said, what we had talked about. [69-year-old female, African American/black, user]

Frequently, users reported using MHAV's secure messaging feature to request prescription reauthorizations. The participants who used secure messaging for this purpose consistently and enthusiastically endorsed it for streamlining the reauthorization process:

I use [MHAV] all the time for my prescriptions. When they start to run out—[when] the refills require authorization—I will shoot off a [secure message] to my doctor's office and they will call the pharmacist and I just go pick it up. They will send me a [secure message] back saying they have sent the prescription and just to pick it up. I think it is really, really great. [66-year-old female, African American/black, user]

Ideas to Improve MyHealthAtVanderbilt's Functionality for Medication Management and Adherence Support

Participants were averse to receiving dose reminders (ie, reminders to take a dose of medication) from MHAV. In general, participants saw value in dose reminder functionality for children, adolescents, and older patients, but did not like the idea of receiving dose reminders via short message service (SMS) text messages, email, or a phone call. Many participants said they do not use SMS text messaging or email on their mobile phones and thought a phone call reminder would be too

intrusive. Those who did use SMS text messaging thought dose reminders would become burdensome and unnecessary:

You know, if I am driving around, I don't want to get a text message or email thing on my phone saying that you know, something about my health. [56-year-old male, Caucasian/white, user]

Others felt email reminders would become annoying and indicated they would just turn off the reminder without taking the medication either because they did not have the medication with them or they would become accustomed to turning off the reminder:

I pretty much take my medicine. I just don't take it the way I should, but I don't know if I would like getting a reminder every day, twice a day, to take my medicine. Too much. I have just—I guess with me working and being at my desk, it's just too much. As I get these little ding-dong bells that pull up on my email...and it would just be overload. [46-year-old female, Caucasian/white, user]

However, participants were enthusiastic about leveraging MHAV to improve medication management and promote adherence through other functionality. We categorized participants' ideas for improving MHAV's medication management functionality into 3 categories: (1) electronically linking MHAV with pharmacies, (2) MHAV alerting providers to patients' fill/refill nonadherence, and (3) using MHAV to help patients understand their medications. Specific ideas are presented under each category.

Electronically Linking MyHealthAtVanderbilt With Pharmacies

Although participants were satisfied with using MHAV to request prescription reauthorizations, they thought that linking MHAV to pharmacies would have several advantages. They wanted (1) MHAV to send proactive refill/reauthorization reminders to providers and patients, (2) MHAV to automatically send patient-initiated prescription medication fill/refill requests to pharmacies, and (3) MHAV to allow patients to request refills and/or reauthorizations for multiple medications at once.

Participants thought MHAV should send proactive reauthorization reminders to providers and/or patients:

It would be nice if [MHAV] kept up with when the [last prescription] was actually written and, 9 months from now, sent me an email that said, "Our records indicate that your prescription is ending [on] the 15th of next month, would you like a renewal? And what pharmacy?" and you could actually correspond back. Because, in my mind, that would reduce what the clinicians have to do on a daily basis, from all the patients needing renewals on prescriptions. Plus, it's going to help me remember and I'm not going to have a lapse in time frame where I'm struggling to try to get my [medications]. [42-year-old female, Caucasian/white, user]

Table 2. Participant comments about MyHealthAtVanderbilt (MHAV) and medication management.

Themes and subthemes	Participant quotes
How portal users use MHAV to manage their medications	
Review medication list	It's really handy to have all that information there—any time you are going to a different doctor that's outside of Vanderbilt. Because I'm asthmatic and I have like a lot of medications that I only take when necessary—I have a sheet of medications, a full sheet! So it's nice because when you go to another doctor that doesn't have access to pull up your medical history, you can go on MyHealth and print it off. So then I can take it with me when I go to a new doctor and [say], "Here, this is what I'm on." And sometimes it's helpful for me because I found that the pharmacy—sometimes the milligrams do not match [what's in MyHealth]. (45-year-old female, Caucasian/white, user)
Use secure messaging to request prescription reauthorizations	Well, the advantage of doing the little [weekly] pill containers is that when I see that the bottle is getting almost empty, I still have a week's worth right there in my pill containers, so then I can MyHealth them right away. Cause I'll look up and go, "I don't have any refills left this time," and I MyHealth them right away and ask them to send a prescription to the pharmacy. (46-year-old male, African American/black, user)
Ideas to improve MHAV functionality for medication management and adherence support	
Send proactive refill/reauthorization reminders to providers and patients	So, I think something system-based, that would notify the doctor that, "Hey, unless something has changed in the medical record, you need to—let's be proactive and have that 90-day [prescription] ready to be called in or sent in or what have you." Because right now [my pharmacy] has to go back to the doctor's office and get that. You know, there is a time lag of a couple of days, but I think if it was proactively done through [MHAV], to where it was notifying the doctor and yet also notifying me...that your prescription is ready to be picked up, I think [it] would be really helpful. (46-year-old male, Hispanic, nonuser)
Automatically send patient-initiated prescription medication fill/refill requests to pharmacies	So you go into MHAV, "I need this refill." You submit it, and it goes to the pharmacy...And it happens every time, I'm standing out there and I'll go [to the pharmacy] at 11 and it's not ready. They close at like 5:30 and I go back at 4. Oh well they haven't called it in yet. Well we are going to be out, you know, out of medicine. I gotta have insulin for tomorrow...What happens is MHAV will send you a message and say your pharmacy's been notified of this refill. [Then] when you [go] to the pharmacy, the pharmacist will argue with you and you'll be like, "You don't have the medication?" and sometimes that's a two to three day deal getting this medication. (42-year-old female, Caucasian/white, user)
Allow patients to request refills and/or reauthorizations for multiple medications at once	That's one of the things that MHAV could do different is to have a link to the pharmacy. So you go in [to MHAV], I need this refill, you check it, submit it and it goes to the pharmacy...you go in [to MHAV] and it says renew/refill prescriptions and you check the box, say I need medicine A and B and you just check A and B. Prescriptions are already on file. I hit submit and it automatically goes to the pharmacy. (45-year-old male, Caucasian/white, user)
Alert providers to patients' fill/refill nonadherence	I think [linking MHAV and the pharmacy] would help. [If] you get a prescription filled for a 90-day supply of Metformin and then 100 days go by and you haven't called for your refill then something is up. You've either been noncompliant and you have way too many pills left or you're dead, or you're in the hospital or something has happened, you know. The pharmacy here doesn't call me and say, "You know what? It's been 100 days since you refilled your Metformin." And I'm thinking, like...I'm sure [another company] would call you to find out, "What's going on? You haven't gotten your refill lately." (58-year-old female ,Caucasian/white, user)
Help patients understand and manage medication side effects	I should be able to go on MHAV and put in Verapamil and pull it up and it tells me, "This is what Verapamil is, what it's prescribed for, if you start noticing these symptoms, tell your doctor." When you change pills, you don't know what is what because you take [a] whole bunch of pills every day...If there were links [on MHAV] where I could have clicked on that medicine and it said, "These are the common side effects. If you have any of these..." Or tell me, "This cough is going to go away in two weeks," or, "This cough is never going to go away—call your provider." (42-year-old female, Caucasian/white, user)
Help patients understand and avoid medication interactions	If I could go into my record...be able to go in there somehow and see all the meds I'm on and have something say, "These meds all work together great," or, "By taking these three different kinds of medicines you might have [an interaction]." And right now, that's not available to me. I don't know [that] unless I sit down with my doctor. If my other doctor puts something on [my medication list], [it's important] that somehow we know that all these drugs are okay together. (46-year-old male, Hispanic, nonuser)

Participants described how linking MHAV with pharmacies could prevent delayed communication between providers and pharmacies. Several participants described problems that occurred because of slow or poor communication between a provider's office and the pharmacy:

> *The problem I had was, like—with my doctor's direction I can request a medication [reauthorization through MHAV] and they may message me back and they will say something like—this happened last time—the nurse wrote back, "Meds called into pharmacy." Now, what am I supposed to think? Meds are called into the pharmacy, right? So, I pick up the phone and I call the pharmacy and say, "Do you have [my prescription] ready?" [They said] "No." I said "Did my doctor's office call in the prescription?" [They said] "No...There is no record." But [the nurse] messaged me and said, "Meds called into the*

pharmacy." Her interpretation is that—without telling me this—before she gets off work she is going to call them into the pharmacy. So I could have hopped up and said, "Oh, my meds are ready, let me drive down and get my meds"...So once you push that button [to send a MHAV request for a reauthorization], it's a chain of events that has to like go to your provider's [office], then the doctor's order [has to go] to the pharmacy and [then] getting [the prescription] ready. [46-year-old male, African American/black, user]

As a result, participants reported going to pick up their prescriptions and finding out the request had not been sent to the pharmacy or the prescription had not been filled as they had expected. Participants also described going through a separate process for each medication when it was due for a refill or reauthorization rather than being able to request that all of their prescriptions be refilled and/or reauthorized at the same time. Participants described how their medications were started at different times and have different supplies (eg, 30 days vs 90 days) and refill amounts (eg, 1 vs 3 refills) resulting in occasions when 1 medication needs to be refilled or reauthorized, but others cannot be requested. As a result, the process of maintaining long-term medications was laborious for those with multiple medications:

For me—because I am on so many different drugs—they run out at different times and I feel like I am always going to the pharmacy, and I hate that. In fact, last week, it was the end of the month and I have like 5 or 6 that were due. I did go without one of my medications for 3 days, which I knew was going to be okay, just so that I could get like 6 of them filled on the same day, so they would run out 30 days later [at the same time]. That is bothersome to me. [46-year-old female, Caucasian/white, user]

There was widespread enthusiasm for adding functionality to MHAV to address these problems. Participants wanted to be able to request refills for several medications at a single time and have MHAV send the information directly to the pharmacy when it was time for the prescriptions to be refilled (Table 2). One participant described how he used MHAV in conjunction with a mail-order prescription refill service. He described how he could order all of his medications at once, regardless of their refill date:

Okay, so you have the pill bar, and you're filling it and you get it all filled and you look in the bottom [of your bottle] and there's only 3 pills left. I gotta go online to [the refill website], [pull] up my account, and it's got a whole list of your medications, you just put a check and it tells you—conceivably you could check all the medicines and it has a column that says, "You cannot be refilled before such and such a date." So if you put a check there, they'll hold that inactive and when the date comes it trips out and they send it automatically...It tells you how many refills you've got left. So during that time, I know that I can go onto MHAV and just send a brief note to my care provider that says, "This script is going to run out in 3 weeks," and they can refill it for me. Just a couple of key

strokes! And Bam! [58-year-old male, Caucasian/white, user]

The other group members responded enthusiastically, and thought MHAV should add similar features and functionality.

Participants described an ideal system that would start with a patient receiving a refill reminder from MHAV, or logging onto MHAV to request multiple medications, and end with a secure message from MHAV telling them their prescription had been filled and was ready to be picked up from their pharmacy. However, they emphasized the importance of having tailored functionality. For example, participants wanted to be able to specify how they are alerted (eg, email, phone call, or SMS text messaging) with refill/reauthorization reminders and where they want their prescriptions filled:

If you could have [MHAV] contact any pharmacy, and set that in there, that would be very helpful. [58 year-old, female, Caucasian/white, user]

MyHealthAtVanderbilt Alerting Providers to Patients' Fill/Refill Nonadherence

By linking MHAV with pharmacies, participants thought MHAV could be leveraged to monitor patients' refill adherence:

Something I just thought about is the pharmacy at Vanderbilt is not tied in with MyHealth, or the EHR. The doctor says, "You need to take this." But there's no record you ever got it filled and you're taking it. [The pharmacies] need to be tied in with [the EHR] so when you pull it up, you can look and see. "This was filled on July 8th for 90 days." That record is [currently] not there. [45-year-old male, Caucasian/white, user]

Participants wanted providers to follow up with them if they were not refilling their medications on time (Table 2).

Using MyHealthAtVanderbilt to Help Patients Understand Their Medications

Participants also suggested adding MHAV functionality to (4) help patients understand and manage medication side effects, and (5) help patients understand and avoid medication interactions. Participants often sent a secure message to their providers to ask about a medication side effect [32], but they felt that many of their questions could be quickly and easily answered by adding functionality to MHAV. Participants also suggested that MHAV alert patients to possible interactions between their prescribed medications that may be missed when they have several doctors prescribing medications for different purposes:

A lot of drug interactions and stuff are well known and are not particularly well managed because you have so many different doctors prescribing at different times. It would be nice if that logic were built in [to MHAV] and you had some degree of confidence that the more common interactions and stuff are being watched by someone, or [MHAV would let you know] with a little alert. [35-year-old male, Caucasian/white, user]

Discussion

Principal Findings

Patient portals represent a technology with the potential to facilitate better care of patients, but virtually no data exist about the real and potential impact of these portals on medication management and adherence. To our knowledge, this is the first study to investigate ways to leverage an existing portal to help patients better manage their medications and, in turn, become more adherent. Users of the MHAV portal (both high and low users) reported using secure messaging to request prescription reauthorizations, and suggested adding portal features and functionality to facilitate medication management and promote adherence. Specifically, participants would like MHAV to be linked with pharmacies to create functionality that (1) alerts providers when long-term prescriptions need to be reauthorized and alerts patients when they need to be refilled, (2) reduces communication problems between providers' offices and pharmacies, (3) allows patients to request multiple medication refills at once, and (4) alerts providers to patients' prescription fill and/or refill nonadherence. Finally, participants would like MHAV to (5) deliver medication information (eg, side effects, other drugs to avoid) in an accessible and user-friendly format. They also emphasized the importance of being able to tailor both how they receive alerts and which pharmacy they want to use to fill or refill prescriptions. Taken together, these suggestions illustrate patients' readiness for additional portal-related features and functionality to support medication management and adherence.

As health care organizations and providers begin to use portals to educate patients about their medications and support them with adherence, there is an increased need for these efforts to be theoretically driven, evidence-based, and patient centered. Behavioral medicine experts have long advocated for the role theory plays in the design and content of health promotion programs [33,34]. The first and most important reason is that programs grounded in empirically derived theories are more effective than those that are not [35]. Second, programs grounded in the theoretical processes that regulate behavior can specify and test the critical assumptions of a program's components to detect exactly why it worked or failed under certain conditions or with certain populations and how it should be improved [33,36]. Both benefits are essential to developing self-care support tools and content within portals that will successfully reduce the personal, social, and economic burden of medication nonadherence. The participants in our study also saw value in integrating currently disjointed systems (ie, portals with pharmacies) to streamline the refill/reauthorization process and to have providers monitor refill adherence. Thus, future research efforts should investigate both the willingness of pharmacies to integrate with patient portal systems to streamline refill and reauthorization processes and of providers to monitor patients' adherence-related activities. Finally, implementing a portal or a medication management module within a portal should go hand in hand with monitoring use and evidence of stakeholder satisfaction, patient adherence, cost-effectiveness, and impact on clinical care and outcomes.

We also explored the types of patients who are and who are not using MHAV and, separately, the relationship between the degree of use and glycemic control. We found that the groups who often have suboptimal glycemic control (ie, African Americans/blacks and individuals with lower SES) [37] were also less likely to have ever used the portal. We also found that, among portal users, more frequent use was associated with better glycemic control. A few recent studies have reported disparities in patient portal use [19,38-40], including 2 studies with diabetes patients [19,40]. Although our methodology, sample size, and portal in question differ from those studies, we also found that adults with diabetes who were African American/black [19,40] or who had less education [40] were less likely to have ever used the MHAV portal. Furthermore, we identified income and insurance disparities in ever using MHAV, which contributes new findings to this literature. Shaw and Ferranti [19] also reported better glycemic control among portal users versus nonusers. In our sample, there was no difference in A1C between portal users and nonusers. Most nonusers (62%) had A1C values less than 7.0%, suggesting these participants may be using other tools to manage their medications and diabetes. We did find an association between more frequent use of a portal and better glycemic control among users only. This finding extends our limited knowledge about the impact of portals on diabetes outcomes [18]. Although differences in glycemic control based on portal use versus nonuse may be spurious (ie, due to the effects associated with education, income, access to computers, or another variable), our finding that more frequent portal use was associated with better glycemic control among participants who had accessed a portal suggests there may be benefits of using a portal that are independent of the contributions of other characteristics of portal users. Future research with larger samples should explore the independent relationship between frequency of portal use and clinical outcomes, adjusting for race and SES.

Limitations

Because we used a mixed-method approach, our quantitative findings are limited by our qualitative sampling procedure. We had limited variability in A1C among nonusers as a whole, and particularly among African American/black nonusers, to be able to tease apart the relationships (or lack thereof) between race, SES, A1C, and portal use versus nonuse. Although portal nonusers appear to be using strategies or tools to achieve optimal glycemic control despite not using a portal, additional research with larger, more diverse samples (ie, in terms of SES and glycemic control) is needed to both replicate this finding and identify what strategies are being used to maintain glycemic control among patients who do not use portals. In addition, we likely oversampled MHAV users by promoting the study over a listserv. All but one of the nonusers were recruited from clinic waiting rooms, whereas MHAV users often contacted us from seeing flyers in clinic waiting rooms or seeing the listserv announcement. Moreover, we found nonusers (most of whom we contacted by phone) were more difficult to reach for scheduling and reminding about focus group sessions than users (most of whom we contacted by email). Second, our cross-sectional design limits our ability to discern causal relationships (eg, we cannot conclude that more MHAV use

XSL•FO

RenderX

improves glycemic control). Third, our study presents participants' perceptions of the frequency of using a patient portal, how they use it, and what added functionality would support medication management and adherence, which may not adequately reflect actual opinions and/or behaviors. Finally, the generalizability of these findings to patient populations with lower SES, other chronic illnesses, and/or those patients using other patient portals is limited. Therefore, we recommend future research explore these issues using different research methodologies with a wide range of patient populations and portal platforms.

Implications for Meaningful Use

Patient portals are increasingly being used to demonstrate meaningful use under the Medicare and Medicaid EHR Incentive Program, which provides financial incentives to providers and hospitals that demonstrate they are implementing EHRs to meaningfully improve patient care [41]. Meaningful use is demonstrated through the achievement of benchmarks including, but not limited to: maintaining an active and correct medication and medication allergy list, identifying patient-specific educational resources and making those resources available to

patients, performing medication reconciliation, automatically tracking medications from order to administration using assistive technologies in conjunction with an electronic medication administration record, and providing evidence of patients' use and engagement with their PHI [41]. As providers and hospitals leverage portals to achieve these benchmarks, it will be important to monitor and learn from portal users and nonusers, understand reasons for nonuse, and identify how to offer medication management and adherence support within portals that meet the needs of patients while also satisfying meaningful use requirements.

Conclusion

We found that patients use portals to manage their health, including their medications (eg, messaging doctors to reauthorize long-term medications), and are enthusiastic about further leveraging these systems to support medication management. Although some portals have included functionality to support medication reconciliation [26], reduce adverse drug events, and improve patient-provider communication regarding medications [20,42], more functionality is needed to maximize medication adherence promotion.

Acknowledgments

Authors' Contributions

Authors CYO and TAE were responsible for the research design. Authors CYO and LSM facilitated focus group discussions, collected and managed data, and conducted the analyses. Author LSM wrote the methods and results, author CYO revised these sections and wrote the introduction and discussion sections, and author KAW wrote the abstract. All authors reviewed/edited all sections of the manuscript, and approved the final version. The corresponding author and guarantor of this manuscript (CYO) takes full responsibility for the work as a whole, including the study design, access to data, and the decision to submit and publish the manuscript.

Conflicts of Interest

None declared.

References

1. Intensive blood-glucose control with sulphonylureas or insulin compared with conventional treatment and risk of complications in patients with type 2 diabetes (UKPDS 33). UK Prospective Diabetes Study (UKPDS) Group. Lancet 1998 Sep 12;352(9131):837-853. [Medline: 9742976]

2. Cramer JA. A systematic review of adherence with medications for diabetes. Diabetes Care 2004 May;27(5):1218-1224. [Medline: 15111553]

3. Odegard PS, Gray SL. Barriers to medication adherence in poorly controlled diabetes mellitus. Diabetes Educ 2008;34(4):692-697. [doi: 10.1177/0145721708320558] [Medline: 18669811]

4. Kim N, Agostini JV, Justice AC. Refill adherence to oral hypoglycemic agents and glycemic control in veterans. Ann Pharmacother 2010 May;44(5):800-808 [FREE Full text] [doi: 10.1345/aph.1M570] [Medline: 20388863]

5. Bosworth HB. Medication adherence. In: Improving Patient Treatment Adherence: A Clinician's Guide. New York: Springer Science + Business Media; 2010:69-95.

6. Inzucchi SE, Bergenstal RM, Buse JB, Diamant M, Ferrannini E, Nauck M, American Diabetes Association (ADA), European Association for the Study of Diabetes (EASD). Management of hyperglycemia in type 2 diabetes: a patient-centered

XSL•FO
RenderX

approach: position statement of the American Diabetes Association (ADA) and the European Association for the Study of Diabetes (EASD). Diabetes Care 2012 Jun;35(6):1364-1379. [doi: 10.2337/dc12-0413] [Medline: 22517736]

7. Rozenfeld Y, Hunt JS, Plauschinat C, Wong KS. Oral antidiabetic medication adherence and glycemic control in managed care. Am J Manag Care 2008 Feb;14(2):71-75 [FREE Full text] [Medline: 18269302]

8. Ho PM, Rumsfeld JS, Masoudi FA, McClure DL, Plomondon ME, Steiner JF, et al. Effect of medication nonadherence on hospitalization and mortality among patients with diabetes mellitus. Arch Intern Med 2006 Sep 25;166(17):1836-1841. [doi: 10.1001/archinte.166.17.1836] [Medline: 17000939]

9. Hong JS, Kang HC. Relationship between oral antihyperglycemic medication adherence and hospitalization, mortality, and healthcare costs in adult ambulatory care patients with type 2 diabetes in South Korea. Med Care 2011 Apr;49(4):378-384. [doi: 10.1097/MLR.0b013e31820292d1] [Medline: 21368684]

10. Wantland DJ, Portillo CJ, Holzemer WL, Slaughter R, McGhee EM. The effectiveness of Web-based vs. non-Web-based interventions: a meta-analysis of behavioral change outcomes. J Med Internet Res 2004 Nov 10;6(4):e40 [FREE Full text] [doi: 10.2196/jmir.6.4.e40] [Medline: 15631964]

11. Linn AJ, Vervloet M, van Dijk L, Smit EG, Van Weert JC. Effects of eHealth interventions on medication adherence: a systematic review of the literature. J Med Internet Res 2011;13(4):e103 [FREE Full text] [doi: 10.2196/jmir.1738] [Medline: 22138112]

12. Misono AS, Cutrona SL, Choudhry NK, Fischer MA, Stedman MR, Liberman JN, et al. Healthcare information technology interventions to improve cardiovascular and diabetes medication adherence. Am J Manag Care 2010 Dec;16(12 Suppl HIT):SP82-SP92 [FREE Full text] [Medline: 21314228]

13. Haggstrom DA, Saleem JJ, Russ AL, Jones J, Russell SA, Chumbler NR. Lessons learned from usability testing of the VA's personal health record. J Am Med Inform Assoc 2011 Dec;18 Suppl 1:i13-i17 [FREE Full text] [doi: 10.1136/amiajnl-2010-000082] [Medline: 21984604]

14. Osborn CY, Rosenbloom ST, Stenner SP, Anders S, Muse S, Johnson KB, et al. MyHealthAtVanderbilt: policies and procedures governing patient portal functionality. J Am Med Inform Assoc 2011 Dec;18 Suppl 1:i18-i23 [FREE Full text] [doi: 10.1136/amiajnl-2011-000184] [Medline: 21807648]

15. Weingart SN, Rind D, Tofias Z, Sands DZ. Who uses the patient internet portal? The PatientSite experience. J Am Med Inform Assoc 2006;13(1):91-95 [FREE Full text] [doi: 10.1197/jamia.M1833] [Medline: 16221943]

16. Zhou YY, Kanter MH, Wang JJ, Garrido T. Improved quality at Kaiser Permanente through e-mail between physicians and patients. Health Aff (Millwood) 2010 Jul;29(7):1370-1375. [doi: 10.1377/hlthaff.2010.0048] [Medline: 20606190]

17. Grant RW, Wald JS, Schnipper JL, Gandhi TK, Poon EG, Orav EJ, et al. Practice-linked online personal health records for type 2 diabetes mellitus: a randomized controlled trial. Arch Intern Med 2008 Sep 8;168(16):1776-1782. [doi: 10.1001/archinte.168.16.1776] [Medline: 18779465]

18. Osborn CY, Mayberry LS, Mulvaney SA, Hess R. Patient web portals to improve diabetes outcomes: a systematic review. Curr Diab Rep 2010 Dec;10(6):422-435 [FREE Full text] [doi: 10.1007/s11892-010-0151-1] [Medline: 20890688]

19. Shaw RJ, Ferranti J. Patient-provider internet portals--patient outcomes and use. Comput Inform Nurs 2011 Dec;29(12):714-8; quiz 719. [doi: 10.1097/NCN.0b013e318224b597] [Medline: 21697705]

20. Schnipper JL, Gandhi TK, Wald JS, Grant RW, Poon EG, Volk LA, et al. Design and implementation of a web-based patient portal linked to an electronic health record designed to improve medication safety: the Patient Gateway medications module. Inform Prim Care 2008;16(2):147-155. [Medline: 18713531]

21. Turvey CL, Zulman DM, Nazi KM, Wakefield BJ, Woods SS, Hogan TP, et al. Transfer of information from personal health records: a survey of veterans using My HealtheVet. Telemed J E Health 2012 Mar;18(2):109-114. [doi: 10.1089/tmj.2011.0109] [Medline: 22304439]

22. Ross SE, Haverhals LM, Main DS, Bull SS, Pratte K, Lin CT. Adoption and use of an online patient portal for diabetes (Diabetes-STAR). AMIA Annu Symp Proc 2006:1080 [FREE Full text] [Medline: 17238699]

23. Koonce TY, Giuse DA, Beauregard JM, Giuse NB. Toward a more informed patient: bridging health care information through an interactive communication portal. J Med Libr Assoc 2007 Jan;95(1):77-81 [FREE Full text] [Medline: 17252070]

24. Rosenbloom ST, Daniels TL, Talbot TR, McClain T, Hennes R, Stenner S, et al. Triaging patients at risk of influenza using a patient portal. J Am Med Inform Assoc 2012;19(4):549-554. [doi: 10.1136/amiajnl-2011-000382] [Medline: 22140208]

25. Allen M, Iezzoni LI, Huang A, Huang L, Leveille SG. Internet-based coaching to improve patient-clinician communication in primary care. AMIA Annu Symp Proc 2007:861. [Medline: 18693963]

26. Staroselsky M, Volk LA, Tsurikova R, Newmark LP, Lippincott M, Litvak I, et al. An effort to improve electronic health record medication list accuracy between visits: patients' and physicians' response. Int J Med Inform 2008 Mar;77(3):153-160. [doi: 10.1016/j.ijmedinf.2007.03.001] [Medline: 17434337]

27. Morgan D. Focus Groups as Qualitative Research. Thousand Oaks, CA: Sage Publications; 1988.

28. Vanderbilt University Medical Center. What is MyHealthAtVanderbilt? URL: http://www.vanderbilthealth.com/main/28542 [accessed 2013-02-25] [WebCite Cache ID 6Ehj65Hxp]

29. Lincoln YS, Guba EG. Naturalistic Inquiry. Newbury Park, CA: Sage Publications; 1985.

30. Patton MQ. Enhancing the quality and credibility of qualitative analysis. Health Serv Res 1999 Dec;34(5 Pt 2):1189-1208 [FREE Full text] [Medline: 10591279]

31. Mayberry LS, Kripalani S, Rothman RL, Osborn CY. Bridging the digital divide in diabetes: family support and implications for health literacy. Diabetes Technol Ther 2011 Oct;13(10):1005-1012 [FREE Full text] [doi: 10.1089/dia.2011.0055] [Medline: 21718098]

32. Wade-Vuturo AE, Mayberry LS, Osborn CY. Secure messaging and diabetes management: experiences and perspectives of patient portal users. J Am Med Inform Assoc 2013 May 1;20(3):519-525. [doi: 10.1136/amiajnl-2012-001253] [Medline: 23242764]

33. Rejeski WJ, Brawley LR, McAuley E, Rapp S. An examination of theory and behavior change in randomized clinical trials. Control Clin Trials 2000 Oct;21(5 Suppl):164S-170S. [Medline: 11018571]

34. Campbell M, Fitzpatrick R, Haines A, Kinmonth AL, Sandercock P, Spiegelhalter D, et al. Framework for design and evaluation of complex interventions to improve health. BMJ 2000 Sep 16;321(7262):694-696 [FREE Full text] [Medline: 10987780]

35. 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]

36. Kraemer HC, Wilson GT, Fairburn CG, Agras WS. Mediators and moderators of treatment effects in randomized clinical trials. Arch Gen Psychiatry 2002 Oct;59(10):877-883. [Medline: 12365874]

37. Duru OK, Gerzoff RB, Selby JV, Brown AF, Ackermann RT, Karter AJ, et al. Identifying risk factors for racial disparities in diabetes outcomes: the translating research into action for diabetes study. Med Care 2009 Jun;47(6):700-706 [FREE Full text] [Medline: 19480090]

38. Goel MS, Brown TL, Williams A, Hasnain-Wynia R, Thompson JA, Baker DW. Disparities in enrollment and use of an electronic patient portal. J Gen Intern Med 2011 Oct;26(10):1112-1116 [FREE Full text] [doi: 10.1007/s11606-011-1728-3] [Medline: 21538166]

39. Sarkar U, Karter AJ, Liu JY, Adler NE, Nguyen R, Lopez A, et al. The literacy divide: health literacy and the use of an internet-based patient portal in an integrated health system-results from the diabetes study of northern California (DISTANCE). J Health Commun 2010;15 Suppl 2:183-196 [FREE Full text] [doi: 10.1080/10810730.2010.499988] [Medline: 20845203]

40. Sarkar U, Karter AJ, Liu JY, Adler NE, Nguyen R, López A, et al. Social disparities in internet patient portal use in diabetes: evidence that the digital divide extends beyond access. J Am Med Inform Assoc 2011 May 1;18(3):318-321 [FREE Full text] [doi: 10.1136/jamia.2010.006015] [Medline: 21262921]

41. Centers for Medicare & Medicaid Services (CMS), HHS. Medicare and Medicaid programs; electronic health record incentive program--stage 2. Fed Regist 2012;77(45):13698-13829 [FREE Full text]

42. Weingart SN, Hamrick HE, Tutkus S, Carbo A, Sands DZ, Tess A, et al. Medication safety messages for patients via the web portal: the MedCheck intervention. Int J Med Inform 2008 Mar;77(3):161-168. [doi: 10.1016/j.ijmedinf.2007.04.007] [Medline: 17581772]

Abbreviations

ADA: American Diabetes Association
EASD: European Association for the Study of Diabetes
EHR: electronic health record
MHAV: MyHealthAtVanderbilt
PHI: personal health information
SES: socioeconomic status
SMS: short message service
T2DM: type 2 diabetes mellitus
TRIAD: Translating Research into Action for Diabetes
UKPDS: UK Prospective Diabetes Study
VUMC: Vanderbilt University Medical Center

XSL•FO
RenderX

XSL•FO

RenderX

Original Paper

Effectiveness of Web-Based Self-Disclosure Peer-to-Peer Support for Weight Loss: Randomized Controlled Trial

Mie Imanaka[1], RD, MPH; Masahiko Ando[2], MD, PhD; Tetsuhisa Kitamura[3], MD, MS, PhD; Takashi Kawamura[1], MD, PhD

[1]Health Service, Kyoto University, Kyoto, Japan

[2]Center for Advanced Medicine and Clinical Research, Hospital, Nagoya University, Nagoya, Japan

[3]Division of Environmental Medicine and Population Sciences, Department of Social and Environmental Medicine, Osaka University Graduate School of Medicine, Osaka, Japan

Corresponding Author:
Masahiko Ando, MD, PhD
Center for Advanced Medicine and Clinical Research
Hospital
Nagoya University
65 Tsurumai-cho
Showa-ku
Nagoya, 466-8560
Japan
Phone: 81 52 744 1953
Fax: 81 52 744 1302
Email: mando@med.nagoya-u.ac.jp

Abstract

Background: Obesity is one of the most common public health problems in the industrialized world as a cause of noncommunicable diseases. Although primarily used for one-on-one communication, email is available for uninterrupted support for weight loss, but little is known about the effects of dietitian group counseling for weight control via the Internet.

Objective: We developed a Web-based self-disclosure health support (WSHS) system for weight loss. This study aims to compare the effect of weight change between those using the WSHS and those using the email health support (EHS).

Methods: This study was designed as an open prospective individual randomized controlled trial. Eligible participants were aged 35 to 65 years with a body mass index (BMI) of ≥25.0 in their latest health examination. Participants were randomly assigned to either the WSHS group or the EHS group. Thirteen registered dietitians under the direction of a principal dietitian each instructed 6 to 8 participants from the respective groups. All participants in the WSHS group could receive nutritional advice and calculate their nutritive intake from a photograph of a meal on their computer screen from the Internet sent to them by their dietitian, receive supervision from the registered dietitian, and view fellow participants' weight changes and lifestyle modifications. In the EHS group, a participant could receive one-on-one nutritional advice and calculate his/her nutritive intake from the photograph of a meal on computer screen sent by email from his/her dietitian, without being able to view fellow participants' status. The follow-up period was 12 weeks for both groups. The primary outcome measure was change in body weight. The secondary outcome measure included changes in BMI and waist circumference. The intergroup comparison of the changes before and after intervention was evaluated using analysis of covariance.

Results: A total of 193 participants were randomly assigned to either the WSHS group (n=97) or the EHS group (n=96). Ten from the WSHS group and 8 from the EHS group dropped out during the study period, and the remaining 87 in the WSHS group and 88 in the EHS group were followed up completely. Weight loss was significantly greater in the WSHS group than in the EHS group (–1.6 kg vs –0.7 kg; adjusted P=.04). However, there were few differences in waist circumference between the 2 groups. (–3.3 cm vs –3.0 cm; adjusted P=.71).

Conclusions: Our newly developed WSHS system using forced self-disclosure had better short-term weight loss results. Further study in a longer-term trial is necessary to determine what effects this type of intervention might have on long-term cardiovascular disease.

XSL·FO
RenderX

KEYWORDS

obesity; weight loss; health support; self-disclosure; email; randomized controlled trial

Introduction

Obesity is one of the most common public health problems in the industrialized world as a cause of noncommunicable diseases, such as ischemic heart disease and diabetes mellitus [1,2]. It has been reported that people are more likely to gain weight when obese persons are around them [3]. In a similar way, behavior modification for weight loss might also transmit to others if a person makes an effort to lose weight. In addition, the necessity to enhance the motivation for weight loss in nutritional counseling has been emphasized [4].

In nutritional counseling for weight loss, face-to-face support that takes into consideration the individual's background and personal characteristics is generally conducted by registered dietitians [5,6]. Recently, emails, which are primarily for one-on-one communication, have been used for weight loss [7-12]. Self-disclosure plays a central role in the development and maintenance of relationships [13], and is also thought to be a critical component in enabling the therapeutic progress [14]. Although writing about experiences of weight loss through blogging as a means of self-disclosure has expanded rapidly recently [15], it is unclear whether forced self-disclosure via the Internet would be actually effective for weight loss.

We developed a Web-based self-disclosure health support (WSHS) system through which participants can receive counseling from a registered dietitian and compare their own changes in weight and lifestyle with those of others. This study aimed to compare the weight loss between the WSHS and the email health support (EHS). Our hypothesis is that weight loss would be greater in the WSHS group than in the EHS group.

Methods

Study Design

This study was an open prospective individual randomized controlled trial (UMIN000009147), carried out from July 2008 through February 2009.

Study Participants

For this study, we recruited participants by mail, contacting clients of the Kyoto University Health Service, Japan, urging them to obtain nutritional counseling for weight loss. Men and women aged 35 to 65 years with a body mass index (BMI) of 25.0 kg/m^2 or more from their latest health examination were eligible. Persons who agreed to participate in our study were invited to an initial face-to-face guidance interview. At this interview, those who had been receiving dietary and exercise therapies, or who could not access Internet or email, or who had a current BMI less than 24.5 kg/m^2 were excluded from our intervention.

Baseline Measurements

At the first guidance interview, we obtained written informed consent and baseline characteristics, such as sex, age, body height and weight, and waist circumference, and established the participants' own target level of weight loss. In addition, we conducted a baseline questionnaire survey on the participants' quality of life (QOL) [16,17]. QOL was measured using the Medical Outcomes Study Short-Form 36 survey (SF-36) [16,17], which is a self-reported measure that assesses 4 separate QOL domains, including general health perception, vitality, role of functioning related to physical and emotional problems, and mental health. Higher scores indicate a more positive health-related QOL for each item [16,17].

Randomization

The participants were randomly assigned to either the WSHS group or the EHS group using the minimization method, balancing sex (male or female), age (<40 years or ≥40 years), and baseline body weight (<60 kg, 60-80 kg, or ≥80 kg) by 1 of the authors (MA). Then they were assigned a counselor-dietitian. A total of 13 registered dietitians under the direction of a principal dietitian provided nutritional counseling. Each dietitian was allocated to both 1 of the WSHS groups and 1 of the EHS groups to minimize the intergroup differences in dietitians' counseling, with 6 to 8 participants of a group being supported by 1 assigned dietitian during the study period. Each dietitian uniformly counseled the participants in both groups based on the standardized manual on the nutritional values of diet records and a photograph of a meal [18] provided by the principal dietitian to maintain the homogeneity of guidance among dietitian counselors. The follow-up period was 12 weeks for both groups.

Interventions and Follow-up

The WSHS group members were given a personal account and password, and could freely access the WSHS system (Figure 1). Each participant set his/her own username and target body weight at the beginning. All members were requested to fill in their present body weight and the level of their lifestyle modification attained such as food records and exercise, along with their motivation level, which were expressed in a 3-level scale (good, fair, and poor), on the screen of the individual's system Web page every week. Participants received nutritional advice and had their nutritive intakes calculated by their dietitians using a photograph of a meal. A participant and his/her dietitian could discuss their questions and comments in this personal area. In this system, group members could view their

fellow participants' weight changes (not actual values) and their related conditions. A participant and his/her dietitian could put their queries or comments on the participant's individual screen, but fellow participants could not write in this column.

The EHS group members were provided with a Microsoft Excel file. They set their target body weight loss at the beginning, and subsequently filled in their present body weight, their levels of lifestyle modification attainment, and their motivation level, similar to the WSHS group members. They could send questions and receive nutritional advice and photo-based nutritive intakes by email. However, the EHS was not a Web-based system, and the participants could not obtain information on their fellow participants' health status via the Web.

The difference between the 2 interventions was that WSHS participants could receive advice from the corresponding dietitian and view other participant's progress when they accessed this system, and EHS participants could only receive advice from the corresponding dietitian. After 12 weeks of online health support, the participants were asked to come in for remeasurement of their height, weight, waist circumference, and QOL by the same dietitian they saw at the beginning of the study.

Statistical Analysis

The primary outcome measure was change in body weight. The secondary outcome measure included changes in BMI, waist circumference, and QOL.

The sample size was calculated based on weight loss during the 12 weeks. We hypothesized that participants assigned to the WSHS group would lose a mean of 2.0 kg after the 12-week intervention, compared with a loss of 1.0 kg in the EHS group with standard deviations of 2.0 kg for both groups [19-21]. Based on 0.9 power to detect a significant difference (P=.05, 2-sided), 85 participants were required for each study group. To compensate for possible absences, we enrolled 90 participants per group.

We conducted intention-to-treat analyses in this study. All data are expressed as mean (SD). BMI was calculated as weight/height^2. Baseline characteristics were compared between the groups using unpaired Student t test for numerical variables and Pearson chi-square test for categorical variables. The comparison of changes before and after intervention between the groups was evaluated using analysis of covariance adjusted for sex, age, and the baseline value of the corresponding item at the first guidance interview. All statistical analyses were performed using JMP 9 statistical software (SAS Institute, Inc, Cary, NC, USA). All tests were 2-tailed and P values of <.05 were considered statistically significant.

Ethical Considerations

All procedures were conducted according to the Declaration of Helsinki. Participants submitted their written informed consent before participation. This study was approved by the Ethics Committee of Kyoto University Graduate School of Medicine.

Figure 1. Screen view of the Web-based self-disclosure (WSHS) health support system.
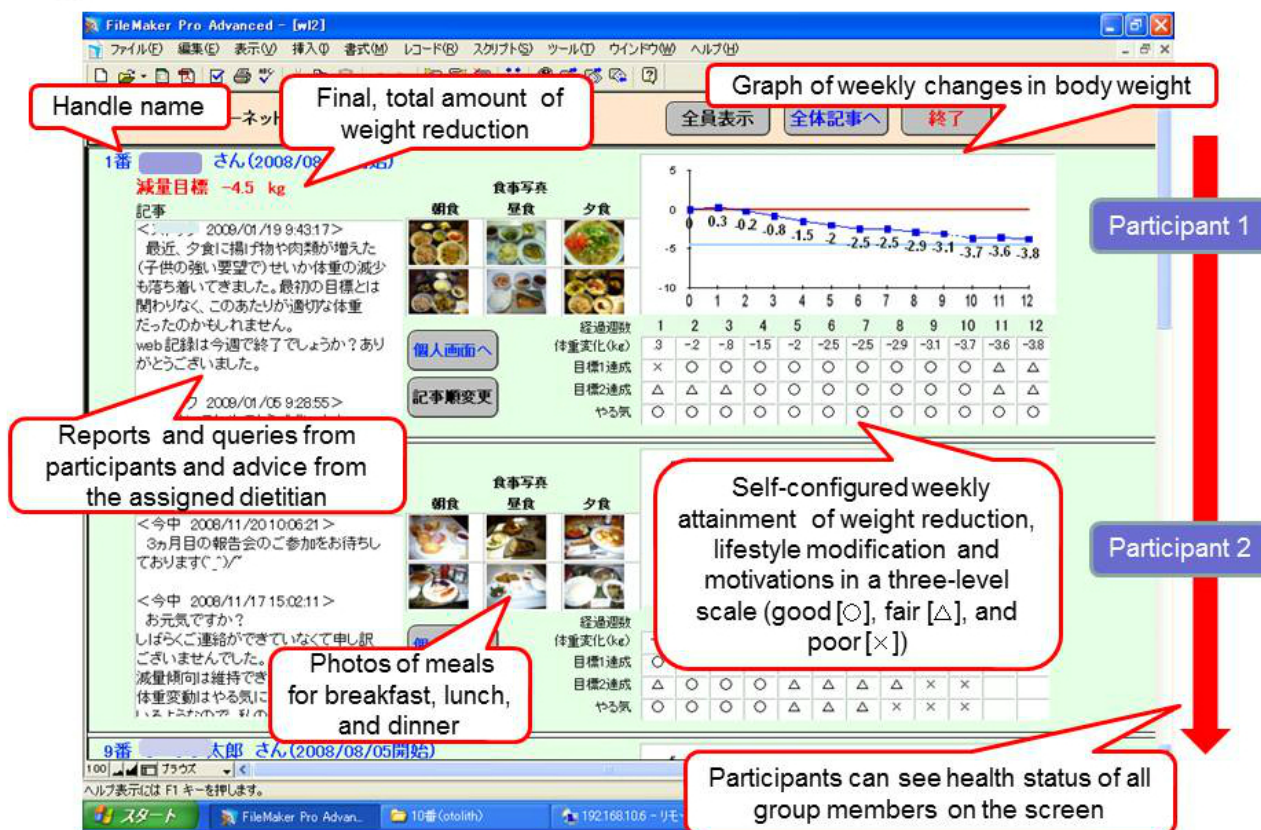
Results

A total of 196 participants were recruited for this trial from July 2008 through February 2009. Among them, 3 participants with a BMI of <24.5 at the first guidance interview were excluded, and the remaining 193 were randomly assigned to either the WSHS group (97) or the EHS group (96). Ten persons from the WSHS group and 8 from the EHS group dropped out during the study period, leaving 87 in the WSHS group and 88 in the EHS group to complete the study (Figure 2).

Baseline characteristics of the participants are shown in Table 1. Mean age was 50 years (SD 7), and mean body weight was approximately 78 kg (SD 10) in both groups. There were no

significant differences in sex ratio, BMI, waist circumference, target body weight loss, or QOL levels between the groups.

Differences in changes of outcomes between the WSHS group and EHS group are shown in Table 2. The loss in body weight was significantly greater in the WSHS group than in the EHS group (–1.6 kg versus –0.7 kg; adjusted P=.04). The decrease in BMI tended to be also greater in the WSHS group than in the EHS group (–0.6 versus –0.3; adjusted P=.05) although it was statistically insignificant. There were no significantly different changes in waist circumference. Changes in QOL scores, general health perception, vitality, role functioning related to physical and emotional problems, and mental health were not significantly different between the groups.

Table 1. Baseline characteristics of participants (N=193).

Participants' characteristics	Web-based self-disclosure health support (n=97)	Email health support (n=96)	P value
Fundamental characteristics			
Male, n (%)	84 (86.6)	82 (85.4)	.81[a]
Age (years), mean (SD)	50.7 (7.4)	49.6 (7.2)	.31[b]
Physical characteristics, mean (SD)			
Body weight (kg)	78.8 (10.6)	77.1 (9.6)	.23[b]
Body mass index (kg/m^2)	27.5 (3.1)	27.4 (2.5)	.28[b]
Waist circumference (cm)	94.9 (7.2)	93.6 (7.2)	.26[b]
Target weight loss (kg)	–4.5 (1.7)	–4.5 (1.9)	.95[b]
Quality of life from SF-36,[c] mean (SD)			
General health perception	52.1 (17.3)	52.7 (13.9)	.79[b]
Vitality	55.0 (15.9)	55.4 (16.3)	.86[b]
Role functioning[d]	67.1 (21.3)	69.9 (17.7)	.34[b]
Mental health	66.0 (17.0)	66.2 (14.6)	.93[b]

[a]Pearson chi-square test.

[b]Student t test.

[c]Data from 86 participants in the Web-based self-disclosure health support group and 88 participants in the email health support group.

[d]Role functioning: role functioning related to physical and emotional problems.

Table 2. Changes in physique and quality of life before and after intervention between groups (N=175).

Changes in outcomes	Web-based self-disclosure health support (n=87)	Email health support (n=88)	P value[a]	
			Unadjusted	Adjusted[b]
Physical changes mean (SD)				
Body weight (kg)	−1.6 (2.7)	−0.7 (2.3)	.02	.04
Body mass index (kg/m^2)	−0.6 (1.0)	−0.3 (0.8)	.03	.05
Waist circumference (cm)	−3.3 (3.3)	−3.0 (3.9)	.64	.71
Changes in items on quality of life from SF-36,[c] mean (SD)				
General health perception	0.4 (13.3)	−1.4 (11.5)	.44	.36
Vitality	0.0 (13.3)	−0.6 (13.3)	.80	.98
Role functioning[d]	4.0 (26.1)	4.7 (19.1)	.89	.70
Mental health	0.5 (17.1)	1.1 (11.1)	.82	.66

[a]Analysis of covariance.

[b]P values were adjusted for sex, age, and the baseline value of the corresponding item at the first guidance interview.

[c]Data from 57 participants in the Web-based self-disclosure health support system group and 59 participants in the email health support group were acquired.

[d]Role functioning: role functioning related to physical and emotional problems.

Figure 2. Patient flow.
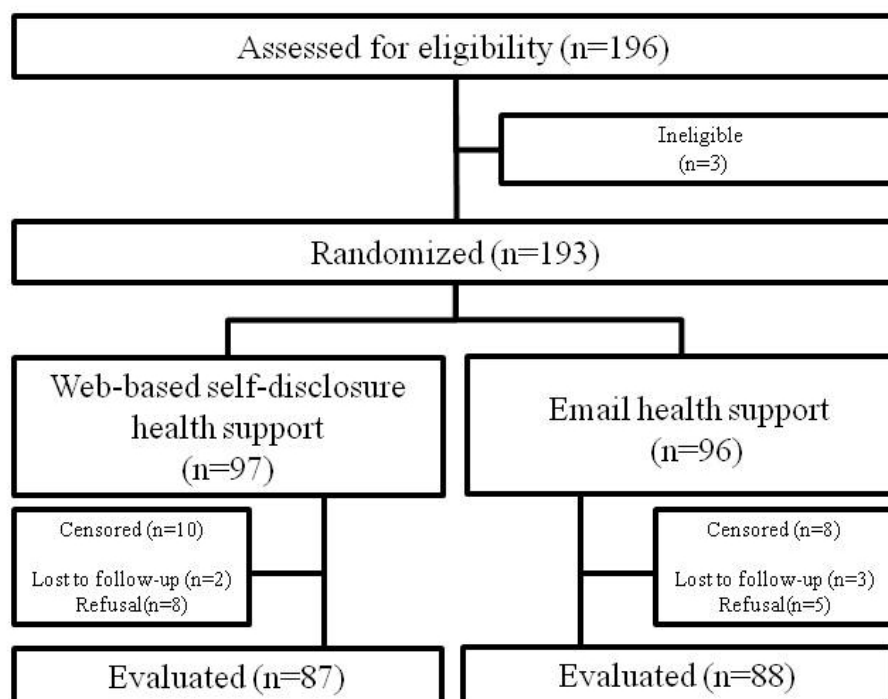
Discussion

Principal Findings

We developed the WSHS system, which is capable of activating self-disclosure, and evaluated its effectiveness in controlling body weight in a randomized controlled trial. In this trial, the WSHS yielded participants with a significantly greater weight loss than did the EHS. Although previous studies developed and evaluated weight loss applications [22-25], our WSHS system is a first step in exploring the benefits of a system in which there is forced self-disclosure and helpful information on nutritional counseling.

It is well-known that face-to-face counseling is effective for weight loss and blood sugar control [5,6]. However, it would be exceedingly difficult to continue face-to-face support for busy middle-aged workers who are candidates for cardiovascular diseases [8,26]. Here, Internet health care support, including email counseling, has shown itself to be similarly effective as

the face-to-face method [7-12]. Because Internet services, including e-learning, are useful [22,23], our WSHS system can also be a useful tool for people who need health support. In addition, the beneficial effects of self-disclosure in health care counseling should be emphasized [4]. The participants were fully aware of their fellow participants' health status, which might have encouraged them to attain their own health goals.

Although reduction in body weight was significantly greater in the WSHS group than in the EHS group, there were no significant differences in waist circumference between the groups. Intra-abdominal fat accumulation of 100 cm^2 or more is one of the cardiovascular risk factors [27]. An 85-cm waist circumference corresponds to approximately 100 cm^2 of intra-abdominal fat accumulation, and is used as a simple diagnostic criterion of metabolic syndrome in Japan [28]. It is known that the measurement of waist circumference varies with participants' intention to flatten their stomach [29], and this flattening might have affected the evaluation of waist circumference in this study.

This study also evaluated 4 SF-36 items as weight loss-associated QOL, but there were few differences in the changes between groups. In previous studies, QOL in obese persons was lower because they were more likely to have low back pain, joint disorders, sleep disorders, and depression [30-32]. Patients with metabolic syndrome were also more likely to be depressive [33]. Furthermore, persons who had succeeded in weight loss improved their own general perception of health and physical functioning [31,34]. The short length of our study or the small intergroup difference in weight loss might have blurred the effects on QOL. Further investigation is needed to evaluate whether WSHS will lead to greater QOL [35]. The number of Internet users reached approximately 2.3 billion people in 2012 across the globe [36]. Therefore, nutritional counseling via the Internet can reach more people and offer more continuous professional support [37]. The WSHS system might cost more to produce than the EHS system; therefore, the cost-effectiveness is an important issue that should be discussed in the future. However, the WSHS systems using forced self-disclosure could be a promising means for decreasing the current high rates of obesity.

Limitations

This study has some inherent limitations. First, the WSHS system is limited to those who can use a personal computer. Second, we only observed the effect on weight loss. The true endpoint of nutritional counseling should be the reduction in mortality and morbidity from cardiovascular diseases. Weight loss could prevent cardiovascular events [38]. Therefore, we consider that weight change is reasonable as a short-term index. Third, lack of information on the number of interactions over the computer/Web interface over the study period and the dietitian interaction with each participant was another limitation because their differences might influence the difference in weight loss among participants. Fourth, information was lacking on social influence on participants' behaviors, which might also have a possible effect on the change of weight loss. Finally, although weight change has been commonly used as a reasonable short-term index, the short observation period of this study is another important limitation because obesity is a chronic condition and requires long-term solutions.

Conclusions

Our newly developed WSHS system using forced self-disclosure would be significantly more effective than the EHS system for short-term weight loss. A longer-term trial that further explores the theoretical differences between these 2 interventions would be necessary to draw conclusions about the WSHS effect on longer-term health conditions.

Conflicts of Interest

None declared.

Multimedia Appendix 1

CONSORT-EHEALTH checklist V1.6.2 [39].

[jmir_v15i7e136_app1.pdf, 1001KB - jmir_v15i7e136_app1.pdf]

References

1. Flegal KM, Carroll MD, Kit BK, Ogden CL. Prevalence of obesity and trends in the distribution of body mass index among US adults, 1999-2010. JAMA 2012 Feb 1;307(5):491-497. [doi: 10.1001/jama.2012.39] [Medline: 22253363]
2. Must A, Spadano J, Coakley EH, Field AE, Colditz G, Dietz WH. The disease burden associated with overweight and obesity. JAMA 1999 Oct 27;282(16):1523-1529. [Medline: 10546691]

3. Christakis NA, Fowler JH. The spread of obesity in a large social network over 32 years. N Engl J Med 2007 Jul 26;357(4):370-379. [doi: 10.1056/NEJMsa066082] [Medline: 17652652]

4. Ministry of Health, Labour and Welfare, Health Service Bureau. Ministry of Health, Labour and Welfare. Program concerning standard specific health diagnosis and health guidance (final version) [in Japanese] URL: http://www.mhlw.go.jp/bunya/kenkou/seikatsu/pdf/02.pdf [accessed 2012-08-20] [WebCite Cache ID 6A3AEdbB6]

5. Dalle Grave R, Calugi S, Centis E, Marzocchi R, El Ghoch M, Marchesini G. Lifestyle modification in the management of the metabolic syndrome: achievements and challenges. Diabetes Metab Syndr Obes 2010;3:373-385 [FREE Full text] [doi: 10.2147/DMSOTT.S13860] [Medline: 21437107]

6. Anderson JW, Reynolds LR, Bush HM, Rinsky JL, Washnock C. Effect of a behavioral/nutritional intervention program on weight loss in obese adults: a randomized controlled trial. Postgrad Med 2011 Sep;123(5):205-213. [doi: 10.3810/pgm.2011.09.2476] [Medline: 21904103]

7. Tate DF, Jackvony EH, Wing RR. Effects of Internet behavioral counseling on weight loss in adults at risk for type 2 diabetes: a randomized trial. JAMA 2003 Apr 9;289(14):1833-1836. [doi: 10.1001/jama.289.14.1833] [Medline: 12684363]

8. Tate DF, Wing RR, Winett RA. Using Internet technology to deliver a behavioral weight loss program. JAMA 2001 Mar 7;285(9):1172-1177. [Medline: 11231746]

9. Collins CE, Morgan PJ, Jones P, Fletcher K, Martin J, Aguiar EJ, et al. Evaluation of a commercial web-based weight loss and weight loss maintenance program in overweight and obese adults: a randomized controlled trial. BMC Public Health 2010;10:669 [FREE Full text] [doi: 10.1186/1471-2458-10-669] [Medline: 21047432]

10. Tate DF, Jackvony EH, Wing RR. A randomized trial comparing human e-mail counseling, computer-automated tailored counseling, and no counseling in an Internet weight loss program. Arch Intern Med 2006;166(15):1620-1625. [doi: 10.1001/archinte.166.15.1620] [Medline: 16908795]

11. Harvey-Berino J, Pintauro S, Buzzell P, Gold EC. Effect of internet support on the long-term maintenance of weight loss. Obes Res 2004 Feb;12(2):320-329. [doi: 10.1038/oby.2004.40] [Medline: 14981225]

12. Appel LJ, Clark JM, Yeh HC, Wang NY, Coughlin JW, Daumit G, et al. Comparative effectiveness of weight-loss interventions in clinical practice. N Engl J Med 2011 Nov 24;365(21):1959-1968. [doi: 10.1056/NEJMoa1108660] [Medline: 22085317]

13. Collins NL, Miller LC. Self-disclosure and liking: a meta-analytic review. Psychol Bull 1994 Nov;116(3):457-475. [Medline: 7809308]

14. Chaudoir SR, Fisher JD. The disclosure processes model: understanding disclosure decision making and postdisclosure outcomes among people living with a concealable stigmatized identity. Psychol Bull 2010 Mar;136(2):236-256 [FREE Full text] [doi: 10.1037/a0018193] [Medline: 20192562]

15. Leggatt-Cook C, Chamberlain K. Blogging for weight loss: personal accountability, writing selves, and the weight-loss blogosphere. Sociol Health Illn 2012 Sep;34(7):963-977. [doi: 10.1111/j.1467-9566.2011.01435.x] [Medline: 22150320]

16. Ware J, Kosinski M, Dewey J. How to Score Version 2 of the SF-36 Health Survey (Standard & Acute Forms). Lincoln, RI: Quality Metric Incorporated; 2001.

17. Fukuhara S, Suzukamo Y. Manual of the SF-36v2 [in Japanese]. Kyoto, Japan: Institute for Health Outcomes & Process Evaluation Research; 2004.

18. Japan Diabetes Society. Food exchange list for diabetic diet therapy 6th edition [in Japanese]. Tokyo, Japan: Japan Diabetes Society; Jul 14, 2012.

19. Committee diagnostic criteria for metabolic syndrome. Definition and diagnostic criteria of metabolic syndrome [in Japanese]. The Journal of the Japanese Society of Internal Medicine 2005;95(4):188-203 http://www.webcitation.org/6HUPn4D81 [FREE Full text]

20. The Japan Society for the Study of Obesity. Committee to create guidelines for obesity treatment. In: 2006 Guidelines for Treatment of Obesity [in Japanese]. Tokyo, Japan: The Japan Society for the Study of Obesity; Apr 15, 2006.

21. Utsugi S, Nagata J, Sato K, Ishizuka K, Kubota A. Effects of a weight loss program that set the behavioral goals to reduce body weight based on the target consumption energy [in Japanese]. URL: http://www.dietitian.or.jp/jdainfo/report/pdf/55-3jissenjirei.pdf [accessed 2012-08-20] [WebCite Cache ID 6A3Bm1ZJ0]

22. Richardson CR, Buis LR, Janney AW, Goodrich DE, Sen A, Hess ML, et al. An online community improves adherence in an internet-mediated walking program. Part 1: results of a randomized controlled trial. J Med Internet Res 2010;12(4):e71 [FREE Full text] [doi: 10.2196/jmir.1338] [Medline: 21169160]

23. Anderson-Bill ES, Winett RA, Wojcik JR. Social cognitive determinants of nutrition and physical activity among web-health users enrolling in an online intervention: the influence of social support, self-efficacy, outcome expectations, and self-regulation. J Med Internet Res 2011;13(1):e28 [FREE Full text] [doi: 10.2196/jmir.1551] [Medline: 21441100]

24. van Genugten L, van Empelen P, Boon B, Borsboom G, Visscher T, Oenema A. Results from an online computer-tailored weight management intervention for overweight adults: randomized controlled trial. J Med Internet Res 2012;14(2):e44 [FREE Full text] [doi: 10.2196/jmir.1901] [Medline: 22417813]

25. Moore TJ, Alsabeeh N, Apovian CM, Murphy MC, Coffman GA, Cullum-Dugan D, et al. Weight, blood pressure, and dietary benefits after 12 months of a Web-based Nutrition Education Program (DASH for health): longitudinal observational study. J Med Internet Res 2008;10(4):e52 [FREE Full text] [doi: 10.2196/jmir.1114] [Medline: 19073541]

26.   Tanaka M, Adachi Y, Adachi K, Sato C. Effects of a non-face-to-face behavioral weight-control program among Japanese overweight males: a randomized controlled trial. Int J Behav Med 2010 Mar;17(1):17-24. [doi: 10.1007/s12529-009-9057-1] [Medline: 19685190]

27.   Hiuge-Shimizu A, Kishida K, Funahashi T, Ishizaka Y, Oka R, Okada M, et al. Absolute value of visceral fat area measured on computed tomography scans and obesity-related cardiovascular risk factors in large-scale Japanese general population (the VACATION-J study). Ann Med 2012 Feb;44(1):82-92. [doi: 10.3109/07853890.2010.526138] [Medline: 20964583]

28.   Examination Committee of Criteria for Obesity Disease in Japan, Japan Society for the Study of Obesity. New criteria for "obesity disease" in Japan. Circ J 2002 Nov;66(11):987-992 http://www.webcitation.org/6HU1rAFuw [FREE Full text]

29.   Imafuku Y. Evaluation of the Measurement of Waist Circumference [in Japanese]. The Official Journal of the Japanese Society of Laboratory Medicine 2007;55(12):1097-1102.

30.   Livingston EH, Ko CY. Use of the health and activities limitation index as a measure of quality of life in obesity. Obes Res 2002 Aug;10(8):824-832. [doi: 10.1038/oby.2002.111] [Medline: 12181392]

31.   Kushner RF, Foster GD. Obesity and quality of life. Nutrition 2000 Oct;16(10):947-952. [Medline: 11054600]

32.   Vetter ML, Wadden TA, Lavenberg J, Moore RH, Volger S, Perez JL, et al. Relation of health-related quality of life to metabolic syndrome, obesity, depression and comorbid illnesses. Int J Obes (Lond) 2011 Aug;35(8):1087-1094 [FREE Full text] [doi: 10.1038/ijo.2010.230] [Medline: 21042326]

33.   Samper-Ternent R, Al Snih S. Obesity in Older Adults: Epidemiology and Implications for Disability and Disease. Rev Clin Gerontol 2012 Feb 1;22(1):10-34 [FREE Full text] [Medline: 22345902]

34.   Patrick DL, Skalicky AM, Edwards TC, Kuniyuki A, Morales LS, Leng M, et al. Weight loss and changes in generic and weight-specific quality of life in obese adolescents. Qual Life Res 2011 Aug;20(6):961-968. [doi: 10.1007/s11136-010-9824-0] [Medline: 21188537]

35.   D'Hondt M, Vanneste S, Pottel H, Devriendt D, Van Rooy F, Vansteenkiste F. Laparoscopic sleeve gastrectomy as a single-stage procedure for the treatment of morbid obesity and the resulting quality of life, resolution of comorbidities, food tolerance, and 6-year weight loss. Surg Endosc 2011 Aug;25(8):2498-2504. [doi: 10.1007/s00464-011-1572-x] [Medline: 21359900]

36.   Internet World Stats. Internet users in the world URL: http://www.internetworldstats.com/stats.htm [accessed 2012-08-20] [WebCite Cache ID 6A3D7qMD0]

37.   Stevens VJ, Funk KL, Brantley PJ, Erlinger TP, Myers VH, Champagne CM, et al. Design and implementation of an interactive website to support long-term maintenance of weight loss. J Med Internet Res 2008;10(1):e1 [FREE Full text] [doi: 10.2196/jmir.931] [Medline: 18244892]

38.   Goodpaster BH, Delany JP, Otto AD, Kuller L, Vockley J, South-Paul JE, et al. Effects of diet and physical activity interventions on weight loss and cardiometabolic risk factors in severely obese adults: a randomized trial. JAMA 2010 Oct 27;304(16):1795-1802 [FREE Full text] [doi: 10.1001/jama.2010.1505] [Medline: 20935337]

39.   Eysenbach G, CONSORT-EHEALTH Group. 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]

Abbreviations

BMI: body mass index
EHS: email health support
QOL: quality of life
WSHS: Web-based self-disclosure health support

XSL•FO

RenderX

Original Paper

An Experimental Test of the Persuasive Effect of Source Similarity in Narrative and Nonnarrative Health Blogs

Amy Shirong Lu[1], PhD

School of Communication, Department of Communication Studies, Northwestern University, Evanson, IL, United States

Corresponding Author:
Amy Shirong Lu, PhD
School of Communication
Department of Communication Studies
Northwestern University
2240 Campus Drive
Evanson, IL, 60208
United States
Phone: 1 (847) 467 4197
Fax: 1 (847) 467 1036
Email: amylu@northwestern.edu

Abstract

Background: Blogs, or websites containing online personal journals, are a form of popular personal communication with immense potential for health promotion.

Objective: Narratives are stories with a beginning, middle, and end that provide information about the characters and plot. Source similarity refers to the degree to which the message source and each recipient are alike with respect to certain attributes. Narratives and source similarity have seldom been examined in tandem as strategies for health persuasion. Personal health blogs provide a suitable platform for such an investigation. This study examined the persuasive effects of message type and source similarity on participants' intentions to adopt a specific health behavior (running for exercise).

Methods: A total of 150 participants were randomly assigned to conditions (n=25 per condition) in a completely crossed, 2 (message type: narrative and nonnarrative) × 3 (source similarity: no similarity, non-health-related similarity, and health-related similarity) between-subjects experiment. First, in an online questionnaire, participants provided personal information in 42 categories and rated the relatedness of each category to running and then completed pretest measures of the dependent variables. Based on their responses, 150 personal health blogs were created. Two weeks later, the initial participants read the blog created with their personal characteristics and completed a questionnaire online.

Results: The source similarity effect was stronger in nonnarrative than narrative blogs. When the blogs were nonnarrative, those with health-related similarities were more persuasive than those with non-health-related similarities. Narrative blogs generated more positive thoughts and stronger blogger identification than nonnarrative blogs.

Conclusions: Health-related source similarity is key for persuasive health communication, especially when the messages are nonnarrative.

Introduction

Background

Almost half of adult American bloggers have posted about health [1]. In addition to affecting readers' health knowledge [2], blogs may also influence their intentions to engage in future healthy behaviors [3]. The existing health blogs could be a rich resource for health behavior promotion for both active participants and lurkers [4].

Some kinds of health blogs may be more effective than others. Suppose a young man has been recently diagnosed with Type 2 diabetes. His doctor says that the diabetes could be managed

if he adopts a healthy diet. The doctor also recommends a smartphone app that, after collecting personal information, automatically retrieves blogs written by other bloggers who are making healthier diet transitions. Suppose the app would select a blog for each reader, would the blog message in the form of personal stories (narrative) or step-by-step instructions (didactic) be more helpful? What kind of blogger would be the most effective in helping this young man achieve his goals? A blogger who uses the same brand of laptop (source similarity not related to the health topic) or a blogger with the same kind of food allergy (source similarity related to the health topic)?

The general research question addressed in this study is what blog features enhance the effectiveness of blogs for delivering persuasive health messages? Two specific blog features were examined. The first feature is whether the blog message is narrative or nonnarrative. A narrative consists of "any cohesive and coherent story with an identifiable beginning, middle, and end that provides information about scene, characters, and conflict, raises unanswered questions or unresolved conflict; and provides resolution" [5], whereas a nonnarrative consists of "arguments, reasoning, claims, and so forth" [6] that are "overtly persuasive messages" [7]. Narratives, or stories, are more available as carriers of personal health information [8-10]. Increasing evidence suggests that narratives are an important persuasion tool with unique capacities [11] and are processed differently than nonnarratives [12].

The second feature is the nature of the blogger's similarity (source similarity) with the user (or blog visitor). Similarity, which refers to the degree to which 2 persons are alike with respect to certain attributes [13], has been found to be an important dimension in persuasion [14-18]. Unrelated similarities may help gain rapport, but not credibility [19]. In this study, similarity refers to matching blogger characteristics to each individual reader. The role of similarity is treated as an empirical question. Specifically, the experimental blogs differed in whether the blogger was depicted as similar to the user in health-related characteristics (or characteristics relevant to the particular health behavior of interest), or as similar in non-health-related characteristics (or characteristics irrelevant to the particular health behavior of interest).

These 2 blog features, the use of narratives and the type of source similarity, were expected to have both independent and joint effects on the users. The interplay of the 2 was empirically explored in personal health blogs as a form of health communication. The specific hypotheses are discussed in the following section.

Hypotheses

Message Type: Narrative vs Nonnarrative

Previous efforts comparing the persuasive effects of narratives and nonnarratives have yielded equivocal findings [20,21]. One possible explanation for such mixed findings is that there have been no standard criteria to compare narratives and nonnarratives. Another important explanation is that many of the studies relied on the dual-processing models of persuasion [22,23], which may not be the most appropriate for studying narratives.

Transportation, along with absorption and engagement, are used interchangeably to indicate people's immersive experience with narratives [24]. A unique mental process, an integrative melding of attention, imagery, and feeling in which a reader becomes absorbed in a narrative world [6], transportation has been identified as a unique persuasive mechanism of narratives as it engages people in behavioral rehearsal [25], appears less overtly persuasive [26], reduces counterargument [27], and enhances message recall [11]. Attitudes formed or changed via transportation are strong, persistent over time, and resistant to counterargument, even though they may not be centrally processed [28]. Blogs, by virtue of their relative personal communication form (almost diary-like) naturally lend themselves to the presentation of personal stories. Therefore, our first hypothesis is that narrative blogs will be more persuasive than nonnarrative blogs.

Source Similarity: Health-Related, Non-Health-Related, None

Similarity, or the degree to which 2 entities are alike with respect to certain attributes [13], appears to influence persuasive outcomes indirectly by affecting the receiver's liking and the receiver's perception of the credibility of the source [29]. Similarity must be studied in conjunction with relatedness [30] because source similarity on unrelated characteristics may help gain rapport, but not credibility [19]. A blog containing non-health-related similarities might plausibly make the blogger more liked because apparent attitudinal similarity increases liking, whereas the health-related similarities might plausibly increase both liking and credibility [29]. Source similarity, therefore, should be about characteristics related to the theme of the message [31,32] to maximize the persuasive outcome. Therefore, our second hypothesis is that (1) the blogs with health-related similarities will be more persuasive than those with non-health-related similarities, (2) the blogs with health-related similarities will be more persuasive than those with no similarities, and (3) the blogs with non-health-related similarities will be more persuasive than those with no similarities.

Interaction of Message Type and Source Similarity

When the blogs are nonnarrative, source similarity should increase personal relevance, elevating the reader's motivation and ability to process the message [22,33]. When the blogs are narrative, source similarity can make it easier for readers to connect or to identify with characters, thus facilitating transportation, which could lead to more persuasion [34]. Therefore, the interaction between message type and source similarity will be treated as an empirical research question: in which message type will source similarity have a stronger persuasion effect?

Methods

Participants

The experimental procedure had been refined by a prior independent pilot study with 30 participants. A total of 300 undergraduate students (150 male and 150 female) from University of North Carolina at Chapel Hill, Chapel Hill, North

XSL·FO
RenderX

Carolina, were invited to participate with a completely cross-balanced, 2 (message type: narrative and nonnarrative) × 3 (source similarity: no similarity, non-health-related similarity, and health-related similarity) between-subjects factorial design. One-half of participants were recruited from 15 classes. One-third were recruited through advertisements posted on the bulletin boards of 3 departments, 3 libraries, and 3 student organizations on campus. The rest were solicited directly from a campus dining hall.

Pretest Questionnaire

The 300 participants were first asked to fill out an online questionnaire, which helped to identify a health behavior most were not doing regularly, but were planning to start or do more regularly; to collect personal information and perceived relatedness of the personal information regarding the health behavior for the subsequent procedures; and to obtain baseline dependent measures.

Three health behaviors (running, yoga, and eating 5 servings of fruits and vegetables per day) were listed as potential options according to recent college health research [35,36]. Each health behavior was feasible to carry out regularly (≥4 times per week) for a typical undergraduate. To ensure that other highly desired healthy behavior was not excluded, participants were also allowed to indicate a behavior they planned to engage in that was not on the list.

Two criteria were used to select the focal health behavior: only a few students should be engaged in the behavior initially, but most should show a moderate level of interest in the behavior in the future. Of the 300 students, 222 (74.0%) completed the pretest questionnaire. Fifty (22.5%) were already regular runners and 185 (83.3%) showed some interest in starting to run or running more regularly. Although 5 (2.3%) were regular yoga practitioners, only 137 (61.7%) expressed interest in practicing yoga in the future. Because 84 (37.8%) participants reported eating 5 servings of fruits and vegetables per day, too few would be eligible for behavior change. Another 61 (27.5%) of the participants mentioned a total of 14 other health behaviors. Thus, of all potential health behaviors, running for exercise best met the selection criteria; therefore, it was used in the main experiment.

Based on extensive reading of running books and magazines and interviews with amateur and professional runners and exercise scientists, 21 characteristics related to running (eg, ideal frequency of running per week) and 21 not related to running (eg, favorite T-shirt color) were identified. Data saturation was reached. Participants were instructed to provide personal information about themselves in all of the 42 categories and rated the relatedness of each category to running based on a 7-point scale (1=extremely unrelated; 7=extremely related). Their answers were carefully screened. If a participant answered questions generally rather than specifically (ie, answered "What is your favorite movie" with "I like many kinds of movies," rather than a specific movie title), the answer was not considered unique personal information and was not used. After all usable answers had been identified, those that had been rated by the participant 4 or greater on the related to running scale were

considered related and those less than 4 were considered unrelated.

Pretest measures of the dependent variables (described subsequently) were also collected.

Blog Prototype Creation

Two single-page blog prototypes (narrative and nonnarrative) were created with the help of 2 professional writers and psychologists and edited to the same length (1293 words). Each prototype either stayed alone as the generic narrative and nonnarrative with no similarity blog entries or allowed the insertion of the blogger's information in up to 42 exclusive categories without invoking logical errors or inconsistencies. However, interviews with another group of 5 undergraduate and graduate students (Question: How many personal characteristics do you think should be embedded into these blog prototypes to make the participants realize the source similarities but not get overwhelmed or suspicious? Answers were coded into numbers and averaged) helped determine that no more than 6 personal characteristics should be embedded in a blog to avoid suspicion and potential negative reaction to the blog. A unique or obscure personal characteristic (eg, "Mack Rice's Three People in Love [1968]" as a favorite song) was inserted as a categorical description (eg, "some early style funky music in the 1960s") instead of a verbatim match to prevent an incredulity reaction. Six personal characteristics related or unrelated to running identified in the pretest were then randomly selected for each participant and were inserted in 1 of the 2 blog prototypes (Figure 1). The no-similarity blogs were adopted directly from the generic prototypes and did not have unique personal characteristics inserted.

The blog was titled "Kerry's online ramblings: A student, a blog, and the life in-between." The blogger was named Kerry to avoid a gender confound because "Kerry" was rated to be the most gender-neutral name among an independent pilot study with 30 participants (mean 2.22, SD 0.71; with 1=male; 2=both; 3=female). As shown in Figure 2, the blog had 3 parts. The left panel showed the blogger's name and basic information. The right panel showed a calendar. The middle panel contained the main experimental blogs.

In the narrative blog prototype, the blogger described how he or she decided to start running (beginning), how difficulties were overcome (development), and ended with a dramatic encounter with a deer during a morning run (climax/ending): "...I ended up choosing an unfamiliar path...Just as I entered the lane and turned a corner...a large deer wandered onto the path. We looked at each other for a brief moment before it nimbly turned around and bounded gracefully off through the woods and disappeared..." In the nonnarrative blog prototype, the blogger provided a total of 15 pieces of didactic suggestions arranged as bullet points on why people should run. For example, the deer encounter appeared as: "Suggestion 12—Feel free to explore new routes when running on trails. New scenery is always helpful, and sometimes you may be lucky enough to encounter some wildlife, such as a deer, like I did."

The 222 participants were randomly assigned to one of the 6 experimental conditions, with 37 persons per condition.

Accordingly, 222 personal blogs were created and uploaded to a Web server. All blogs were edited to be approximately the same length. Design and layout (eg, blogger name, basic information, calendar) were identical across conditions except for manipulation of the 2 independent variables (message type and source similarity) in the main experimental blog section. Each participant saw 1 blog post for this study. The commenting function was disabled for control purpose.
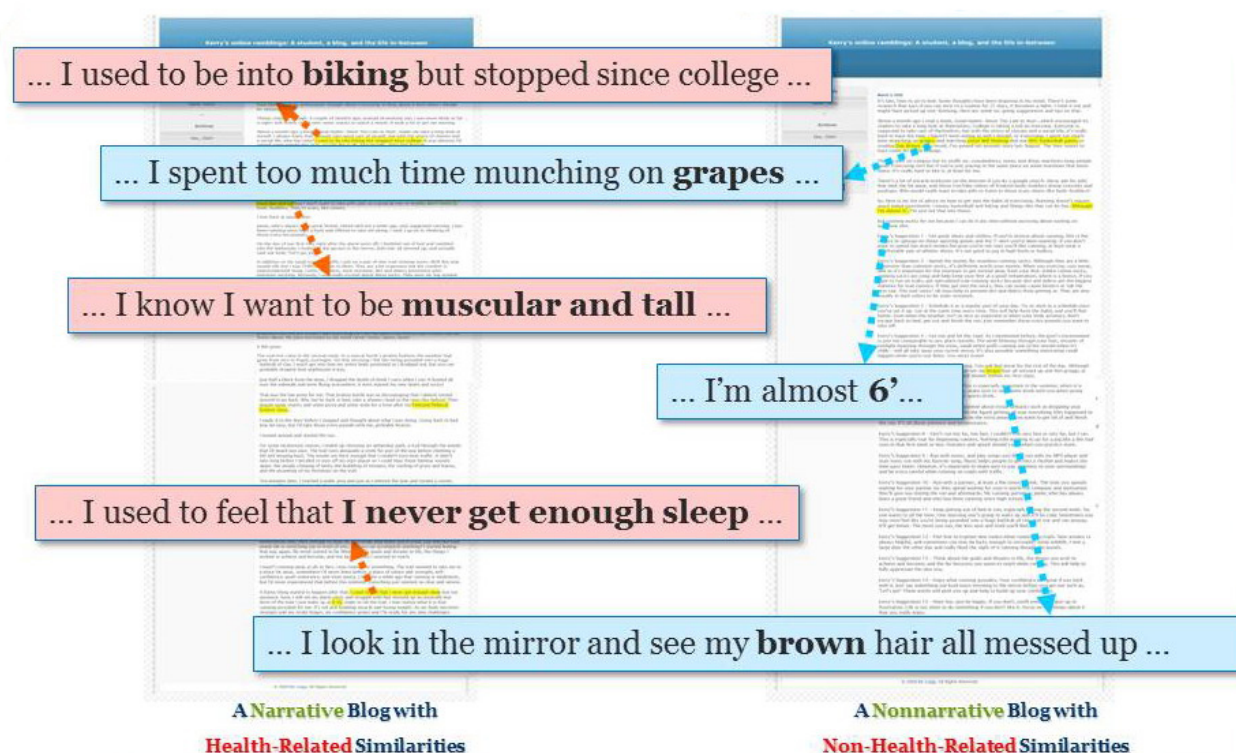
Main Experimental Procedures

At least 2 weeks after the pretest, students participated in the experimental session. To weaken the association between the pretests and posttests, the 2 sessions were promoted with different titles and the questionnaires had different layout designs. In the experiment session, each participant sat at a computer and was instructed to read and sign the consent form. Participants in the health-related similarity and non-health-related similarity conditions saw a (narrative or nonnarrative) blog matched on his or her earlier pretest responses and those in the no-similarity condition saw a generic (narrative or nonnarrative) blog. The time each participant spent reading the blog was recorded by an embedded Web app.

Of the 222 participants who participated in the pretest, 204 (91.9%) completed the posttest. Of the 204 participants, 39 (19%) identified themselves as regular runners who ran at least 30 minutes 4 times a week; 8 (4%) figured out that their answers to the pretest were used to create the blogs and guessed both projects were related; 3 (2%) were nonnative English speakers; 3 (2%) wrote explicitly that they did not want to run at all because of health conditions (depression, fracture, and paralysis); and 1 (0.5%) accidentally completed the survey assigned to another student by sitting at the wrong computer. These 54 (27%) students were excluded from the analysis, resulting in a sample of 150 participants, with 25 participants per condition. Figure 3 provides a flowchart for the complete experimental procedure.

Figure 1. Blog construction.

Figure 2. Experimental blog.

Figure 3. Experimental procedure.
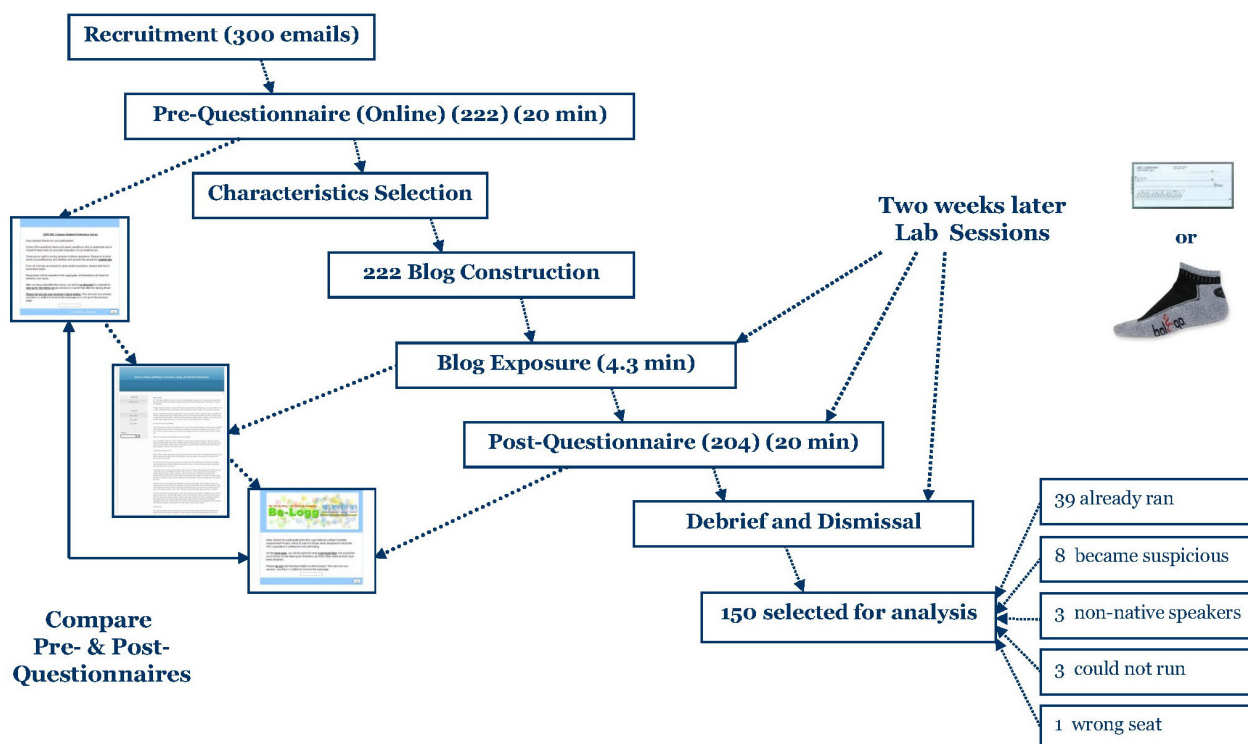
Dependent Measures

The primary outcome variables were 3 independent measures assessing participants' intention to run for exercise based on established measures and consultation with runners and exercise scientists. One was a single-item measure asking participants the likelihood of running in the near future on a unipolar scale [37]: "Please indicate the likelihood you would start running for 30+ minutes for 4+ times per week in the near future (1=not at all likely; 7=extremely likely)." The second asked about the participant's intended running duration: "If you start running in the near future, how long would you like to do it every time? __ Minutes." The third was a behavioral measure: running-related gift selection. Before leaving the computer laboratory, each participant was offered 2 compensation options: a US $10 check or 2 pairs of running socks. Both options were perceived by a separate group of students as worth the same monetary value. The socks were unisex, of different sizes, popular among runners, specifically designed for trail running, and available only in professional running stores. Because the blog advocated running for exercise and the blogger had been running on a trail, participants' choosing a form of compensation specifically designed for the advocated health behavior could serve as an additional indicator of the context-specific persuasion outcome.

Psychological States as Potential Mediating Variables

The main goal was to explore how message type and source similarity affected people's response to persuasive messages; therefore, hypotheses were tested by relying on the blog's intrinsic features and manipulation checks were not conducted [38]. Three variables (narrative transportation, source similarity, and relatedness) were assessed with Likert-style 7-point scales

(1=strongly disagree; 7=strongly agree) and were summed and averaged as participants' psychological states in response to the message manipulation. Narrative transportation was assessed by using Green and Brock's [12] 11-item scale, which demonstrated good internal consistency (α=.88). Sample items included: (1) I want to learn how the blog would end, and (2) I can picture myself in the blog. Source similarity was assessed with a 2-item measure adapted from customization research [39], which exhibited a strong correlation (r=0.80, P<.001): (1) this blog targeted me as a unique individual, and (2) this blog was personalized according to my interests. Relatedness was assessed with a 2-item measure: "The similarities between Kerry and me are…" and "the characteristics shared by Kerry and me are..." related to whether I will begin running 30+ minutes for 4+ times per week in the near future. The 2 items were strongly related (r=0.77, P<.001).

Blog Processing and Involvement Variables

Six theoretically relevant variables were explored. Two assessed readers' blog processing (number of meaning units and thought valence). Meaning units refers to a collection of words related to 1 central meaning, also known as an idea unit [40]. In the posttest questionnaire, participants were asked to list all of the thoughts they had while reading the blog [41]. Two native-English speakers unaware of the experimental purpose served as coders to count the number of relevant meaning units. The coders also coded the valence of each relevant thought (-1 for negative; 0 for neutral; and +1 for positive). The valence scores of each thought were aggregated into a valence scale for each participant.

Four variables addressed readers' involvement with the blogger. All were assessed on 7-point Likert scales (1=strongly disagree; 7=strongly disagree). Perceived source credibility was assessed

with a 6-item scale adapted from Metzger and colleagues [42] and showed good internal consistency (α=.93) (sample items are "I trust the blog Kerry posted" and "Kerry is credible"). Interpersonal attraction was assessed with a 10-item scale adapted from McCroskey and McCain [43] that achieved adequate internal consistency (α=.78) (sample items are "I think Kerry could be a friend of mine" and "I would like to have a friendly chat with Kerry"). Identification with the blogger was measured with 7 items selected from Eyal and Rubin's [44] 10-item scale that achieved strong internal consistency (α=.90) (sample items are "when I was reading the blog, I imagined myself doing the same things Kerry was doing" and "I really felt as if I was Kerry who was running"). Parasocial interaction was assessed with Rubin's [45] scale (α=.90) (sample items are "Kerry makes me feel comfortable, like I'm with a friend" and "I look forward to reading Kerry's blog when more are posted"). Although formal hypotheses were not proposed for the involvement variables, they proved valuable in understanding the process and consequence of the interaction of the message type and source similarity.

Results

Preliminary Analyses

The 150 students came from 33 academic majors across campus (17.3% freshmen, 31.3% sophomores, 20% juniors, and 31.3% seniors). Sixty-eight (45.3%) were male and 82 (55.7%) were female. The mean age was 21.2 years (SD 2.3). They spent 4.3 hours (SD 1.6) online per day. The 150 blogs were carefully edited to be approximately 1300 words across conditions (mean 1316, SD 22). The health-related similarity blogs (mean 1342, SD 12) were longer than the non-related similarity blogs (mean 1314, SD 9), which were longer than the no-similarity, or generic blogs (mean 1293, SD 0; P<.001). There were no significant word count differences between the narrative and nonnarrative blogs. The average reading time was 256.8 seconds (SD 100.7), which did not differ across conditions (P=.20).

Analyses of variance (ANOVAs) were conducted on both likelihood of running and intended running duration by message type and source similarities. Participants across conditions did not differ on the pretest measures, and there was no significant gender or age difference. Paired sample t tests indicated that both posttest measures had significantly increased from the pretest (P<.001).

Although each blog with similar personal characteristics (health-related and non-health-related) differed in specific personal characteristics depending on each individual's ratings, according to the cumulative relatedness ratings across the 42 categories, 5 characteristics, including free time, exercise frequency, ideal body shape, lack of sleep, and current exercise were rated as the overall most related, whereas favorite bookstore, hair color, favorite book, favorite T-shirt color, and favorite news site were rated as the overall least related.

A 2 (narrative and nonnarrative) $\times$ 3 (no similarity, non-health-related similarity, and health-related similarity) ANOVA indicated that people responded differently to narrative and nonnarrative blogs ($F_{1,144}$=13.46, P<.001, partial η^2=0.09).

Transportation was significantly higher in the narrative conditions (mean 4.58, SD 0.96) than in the nonnarrative conditions (mean 3.93, SD 1.19). Neither the main effect or interaction for source similarity on transportation was significant.

Two 2-way ANOVAs were conducted to explore participants' response to source similarity. The first 2 (narrative and nonnarrative) $\times$ 3 (no similarity, non-health-related similarity, and health-related similarity) ANOVA showed a significant main effect for the source similarity ($F_{1,144}$=7.69, P=.001, partial η^2= 0.10). Post hoc Tukey honestly significant difference (HSD) comparisons indicated that the mean scores in health-related (mean 3.84, SD 1.75) and non-health-related similarity conditions (mean 4.00, SD 1.78) were significantly higher on the source similarity scale than those with no similarities (mean 2.77, SD 1.55; P=.001). There was no significant difference between the 2 similarity conditions (P=.89) or any significant main effect or interaction for message type on source similarity type ($F_{1,144}$=1.06, P=.31). The second 2 (non-health-related similarity and health-related similarity) $\times$ 2 (narrative and nonnarrative) ANOVA revealed a significant main effect on relatedness for both similarity conditions ($F_{1,96}$=4.98, P=.03, partial η^2=.05). The relatedness score of participants in health-related similarity (mean 4.18, SD 1.12) was significantly higher than those of participants in the non-health-related similarity (mean 3.66, SD 1.19; P=.001). Neither the main effect nor the interaction for message type on relatedness were significant.

Testing of Hypotheses

To test the hypotheses, two 2 (message type: narrative and nonnarrative) $\times$ 3 (source similarity: no similarity, non-health-related similarity, and health-related similarity) analyses of covariance (ANCOVAs) were performed on the likelihood of running and the intended running duration with the posttest measures as dependent variables and the pretest measures as covariates. In all ANCOVAs, the pretest measures were significant and were not reported individually. For health-related gift selection, the only categorical outcome, full and conditional cross-tabulations were performed.

The first hypothesis predicted that narrative blogs would be more persuasive than nonnarrative blogs. This was not supported by the likelihood of running or the health-related gift selection (chi-square [χ^2_1] =1.43, P=.23). The 2 $\times$ 3 ANCOVAs only revealed a non-significant effect for the intended running duration ($F_{1,143}$=3.30, P=.07, partial η^2=0.02) for narratives (mean 32.60, SD 10.57) over nonnarratives (mean 29.27, SE 9.10). Therefore, the first hypothesis was not supported.

The second hypothesis predicted that blogs with health-related similarities would be more persuasive than those with non-health-related similarities and those with no similarities. For the likelihood of running, source similarity was not significant (P=.15). Post hoc comparisons showed no differences in source similarity across the narrative conditions (P=.98). In the nonnarrative conditions, however, the non-health-related similarity condition (mean 3.04, SD 1.51) was significantly

lower than that of the health-related similarity condition (mean 4.32, SD 1.60; P=.01). Although the non-health-related similarity condition was lower than that of the no similarity (mean 4.12, SD=1.56; P=.06), it did not meet statistical significance. There was no difference between the related and the no-similarity conditions (P=.49). See Figure 4 for the change scores across conditions.

For intended running duration, ANCOVA showed a significant effect for source similarity ($F_{2,143}$=3.17, P=.04, partial η^2=0.04). Post hoc comparisons indicated a similar pattern: the effect was also primarily due to the non-health-related similarity condition (mean 25, SD 8.17), which was lower than the no-similarity generic condition (mean 29.80, SD 8.48; P=.05) and significantly lower than the health-related similarity condition (mean 33, SD 9.13; P=.002). The no-similarity condition generated a shorter intended running duration than the health-related similarity condition, although it was not statistically significant (P=.09). See Figure 5 for the change scores across conditions.

For health-related gift selection, 32 (21.3%) participants chose to get the socks instead of the check. The chi-square test was not significant (χ^2_2=2.23, P=.33). Analyses of the nonnarrative data showed that more participants in the health-related similarity conditions chose the socks (n=10) than the non-health-related (n=3) and the no-similarity conditions (n=6), (P=.02 and P=.04, respectively). See Figure 6 for the scores across conditions.

Thus, the first part of the second hypothesis, the blogs with health-related similarities will be more persuasive than the blogs with non-health-related similarities, was supported when the message was nonnarrative. The second part of the second hypothesis, the blogs with health-related similarities will be more persuasive than the blogs with no similarities, was partially supported by health-related gift selection when the message was nonnarrative. The first hypothesis, the narrative blogs will be more persuasive than nonnarrative blogs, and the third part of the second hypothesis, the blogs with non-health-related similarities will be more persuasive than blogs with no similarities, were not supported.

To answer the research question, in which message type will source similarity have a stronger persuasion effect, the factor of message type (2) × source similarity (3) was examined and found to be not significant for the likelihood of running and the intended running duration (P=.17 and P=.73, respectively). A 2 (narrative and nonnarrative) × 2 (non-health-related similarity and health-related similarity) ANCOVA, however, provided some evidence: the difference between health-related and non-health-related similarity for nonnarratives was bigger than that for narratives ($F_{1,95}$=3.38, P=.06, partial η^2= 0.03) for the likelihood of running. In other words, the effect of whether the similarity was health-related was more pronounced in the nonnarrative conditions than in the narrative conditions. This result, however, did not meet statistical significance. For the health-related gift selection, the differences appeared within source similarity types under nonnarrative conditions (see the second hypothesis and Figure 6), but not in narrative conditions (χ^2_1<0.2). So source similarity seems to have a stronger persuasive effect in the nonnarrative context than in the narrative context.

Figure 4. Interaction of message type and source similarity on likelihood of running change.
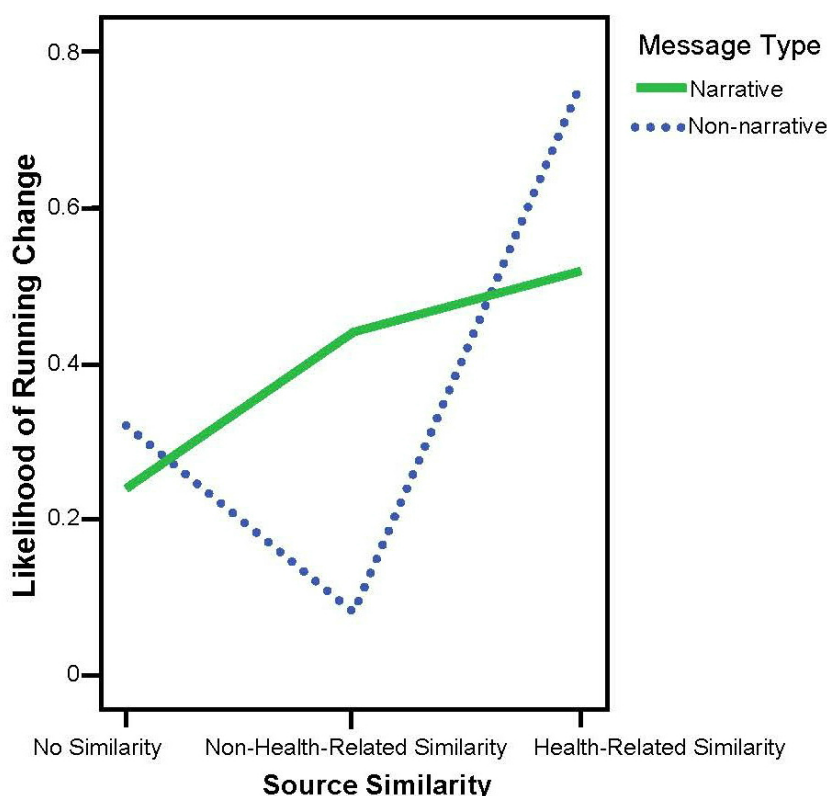
Figure 5. Interaction of message type and source similarity on intended running duration change.
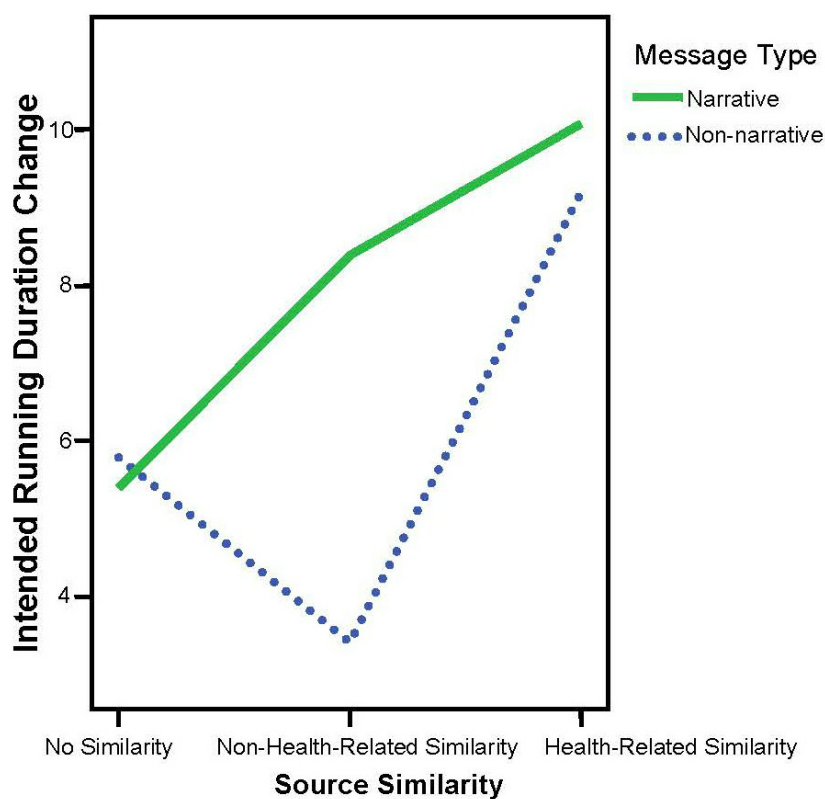
Figure 6. Interaction of message type and source similarity on health-related gift selection.
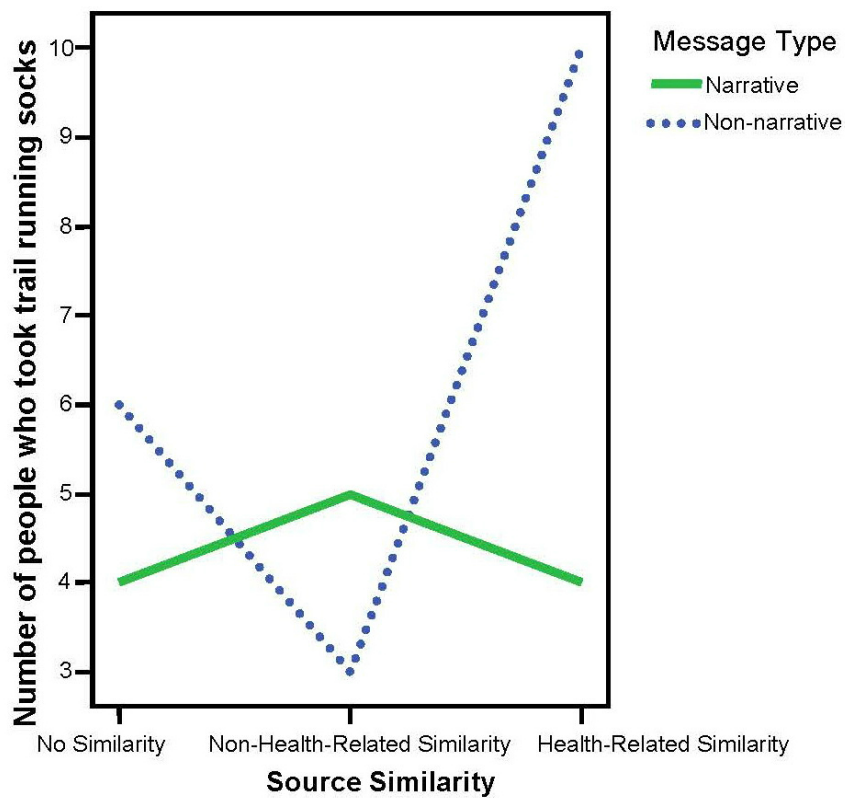
Blog Processing Variables

Analyses of the number of meaning units and thought valence suggested that readers processed narratives and nonnarratives differently. Although the 2 (narrative and nonnarrative) $\times$ 2 (no similarity and similarity including non-health-related and health-related combined) ANOVA was not significant, the post hoc multiple tests showed that for blogs with source similarities, health-related or not, nonnarratives (mean 10.50, SD 6.17) produced more thoughts than narratives (mean 8.52, SD 4.32, P=.06) although the result did not meet statistical significance. Readers also had more positive thoughts about narratives (mean 1.10, SD 4.27) than nonnarratives (mean –0.89, SD 5.53; $F_{1,144}$=6.18; P=.01; partial η^2=0.041).

Involvement Measures

All involvement measures were analyzed by 2 (narrative and nonnarrative) $\times$ 3 (no similarity, non-health-related similarity, and health-related similarity) ANOVAs except for interpersonal attraction, which was analyzed in a 2 (narrative and nonnarrative) $\times$ 2 (no similarity and similarity including both health-related and non-health-related) ANOVA. There was a main effect for source similarity, $F_{1,144}$=3.89, P=.05, partial η^2= 0.03, for perceived source credibility: bloggers with non-health-related similarities were perceived as less credible (mean 4.70, SD 1.18) than bloggers with health-related similarities (mean 5.13, SD 0.65). Source similarity had an effect ($F_{1,146}$=3.22, P=.07, partial η^2= 0.02) on interpersonal attraction: blogs with similarities (mean 4.60, SD 0.67), health-related or non-health-related, were perceived as more attractive than bloggers with no similarities (mean 4.42, SD 0.69) although this did not meet statistical significance.

Readers showed more identification with bloggers of narrative messages (mean 4.20, SD 1.38) than of nonnarrative messages (mean 3.70, SD 1.40; $F_{1,144}$=7.43, P=.007, partial η^2= 0.05). They also showed a higher level of parasocial interaction with bloggers of narrative messages (mean 4.43, SD 1.23) than of nonnarrative messages (mean 4.14, SD 1.29; $F_{1,144}$=4.7; P=.03; partial η^2= 0.03).

Each of the potential mediating and involvement variables was included as a control in the hypotheses testing 2 $\times$ 3 ANCOVA models. Although they were significantly correlated with the outcome variables, and none of the previous significant effects were eliminated, they were not significant in the models, suggesting that these variables were not mediating the effect of message type and source similarity on the dependent variables. Gender was included as a third independent variable in all of the hypotheses testing ANCOVA models with no significant main effects or interactions.

In summary, the source similarity made much more difference in nonnarrative messages than in narrative messages. Nonnarrative blogs were much more persuasive if they contained health-related similarities than if they contain non-health-related similarities. When the blogs were nonnarrative, source similarity led to an increase in the number of thought meaning units. Although bloggers with similarities were perceived to be more attractive than bloggers with no similarities, those with health-related similarities were perceived to be more credible than those with non-health-related similarities. Compared with nonnarrative blogs, narrative blogs elicited more positive thoughts. Readers of narrative blogs identified more with the bloggers and were more likely to feel some parasocial interaction with them than readers of nonnarrative blogs. The source similarity, however, does not make much of a difference to the persuasive outcome narrative blogs.

Discussion

Principal Findings

This study is one of the first systematic empirical examinations of the interplay of message type and source similarity in personal health blogs. Significant differences in healthy behavior intention were detected depending on the kind of source similarity in the nonnarrative health blogs, but not in the narrative health blogs: blogs with health-related similarities were more persuasive than blogs with non-health-related similarities.

Why would source similarity have a stronger persuasion effect in nonnarrative messages than in narrative messages? In the narrative context, although source similarity may make it easier for readers to connect with characters, thus facilitating transportation [46], source similarity, however, is not necessary for transportation. A skilled writer may also make a story relevant to his readers through other elements, such as vivid descriptions and engaging plots [24]. Therefore, when blogs are engaging narratives, readers may be already involved and may not scrutinize the messages or evaluate the relatedness of source similarities as much as those nonnarrative blog readers.

In a nonnarrative context, the blogger's credibility could play more of a role. Therefore, variables that would influence credibility (eg, health-related similarities) will correspondingly have a larger role to play in nonnarratives as compared to narratives. In the nonnarrative context, the personal health blog is more prescriptive than descriptive, which is highly personally relevant. The source similarity manipulation is both wide and deep, which has been confirmed by each recipient's input [47-49]. Readers were less likely to treat the similarities as heuristic cues. Such personal relevance would increase readers' motivation to process persuasive information [22] systematically rather than heuristically [33]. Some evidence existed for this explanation in that readers in both source similarity conditions had significantly more meaningful thought units than those in the no-similarity conditions. As a result of central processing, readers of nonnarrative blogs may have scrutinized the blogs with more sensitivity to the similar personal characteristics than readers of narrative messages, who were already more transported into the story.

Although the source similarity effect was stronger in the nonnarrative conditions, narratives should not be written off as unimportant. The narratives could motivate people to increase their physical activity even when the blogs were not specifically created with source similarities inserted as blogger characteristics. The narrative readers who enjoyed the running experience in the stories might imagine running longer because

of their pleasant reading experience [50]. The positive thought valence for the narratives provides some support for this idea. In contrast, the nonnarratives provided only didactic instructions, which might provide a less enjoyable reading experience. The participants could have felt they were being "talked at," which generated less positive feelings. This pattern could also be explained by psychological reactance theory [51], which posits that reactance will occur when people perceive that their freedom of choice is threatened. People may feel less threatened by narratives and, thus, less resistant to the messages as they are free to use their imaginations as they are transported into the story. In this study, the message type (narrative and nonnarrative) manipulation was based on 2 prototypes previously created without consulting with each participant's narrative/nonnarrative preference, whereas the source similarity (no similarity, non-health-related similarity, health-related similarity) was manipulated thoroughly based on each participant's input. If the message types could be manipulated in the same way, it is likely that a more significant difference could be observed. In fact, although the blogger in the nonnarrative blog conditions was giving a list of suggestions, the blogs were still written in the first person perspective. It also mentioned a personal experience (deer encounter), a conflict (to choose a new route or not), and a solution (yes, to choose a new route), all of which might even evoke some slight unconscious narrative processing among participants as people have an innate tendency to process information as narratives. This could also explain the lack of persuasive outcome difference between the 2 conditions.

This study demonstrated that to be effective, the source similarity "match" must be well integrated with the persuasive messages. The difference between health-related and non-health-related similarities could also be rephrased as the distance between different aspects of the self [4]. An examination of the top 5 most health-related and non-health-related similarities suggested such a pattern: most of the health-related similarities were personal characteristics (eg, exercise frequency) and the non-health-related similarities were personal preferences (eg, favorite book). Personal preferences define what a person likes whereas personal characteristics define who a person is. Therefore, source similarity should be aligned with characteristics that are central to the recipient's self-concept.

Useful insights for health communication can be drawn. Blogs may be a useful tool for encouraging people to adopt healthy habits. In this study, just 1 exposure significantly increased both measures of behavioral intentions. Such interventions could be further integrated with social networking sites. After people provide information about themselves, they could be directed to the online virtual blogger communities that would be most helpful for them. Health communicators should be careful in selecting appropriate personal characteristics on which to create persuasive messages with similar sources. In this study, non-related source similarities could backfire and even cancel out the increased persuasion effect due to related source similarities, resulting in the lack of difference between the similarity and no-similarity conditions.

Health communicators should be extremely careful selecting appropriate characteristics on which to create customized health messages. When resources are too limited to effectively evaluate the most appropriate characteristics for customization, creating a generic transporting narrative message may be the most cost-effective solution.

In fact, a message with both narrative and nonnarrative elements may be most effective. Narratives may help reduce the initial psychological reactance to the persuasion by transporting readers into the narrative world [27]. Once the audience is on board, an appropriately created nonnarrative message with their prior input could be delivered. When done well, such practice should create the feeling of interpersonal communication [52].

Limitations

This study has limitations. The narratives created by academic researchers are usually not as transporting as those written by professional writers. The narrative manipulation could have been more thorough and in-depth. Although the nonnarrative blogs were less prose-like and did not fully conform to the narrative definition [5], they could still retain some narrative elements or even evoke narrative processing. The all-student sample read only 1 experimental blog, and posttest measures were taken immediately after the exposure. Due to the lack of statistical power, results of several statistical comparisons did not meet statistical significance and mediation analyses were precluded. Although choosing the trail running socks might indicate the intention to run, choosing the cash prize did not necessarily indicate an absence of running intention as people may prefer the cash over the socks despite high intentions to exercise. The mix of domains (money vs a specific item) might pose potential threat to internal validity. Other than the gift selection behavior, no actual health behavior was measured. Although no student was able to correctly identify the purpose of the study, demand characteristics could be still at play. For experimental control, participants were unable to post comments on the blogs. Repeated exposure to multiple blog entries, delayed posttest measures, and user comments may result in different attitudinal and behavioral change patterns. The self-concept is highly volatile and can be easily changed by priming [53]. Health professionals could adapt to the volatility of self-concept by identifying more intrinsic and stable characteristics. The pretest questionnaires required participants to answer more than 130 questions for at least 20 minutes. Online users may be reluctant to devote so much time to providing so much personal information. Alternative plans should be devised to collect personal information without arousing suspicion or fatigue.

Conclusions and Future Research

Future research could follow several different paths. Source similarity could and should be treated in a more fine-tuned manner. Instead of emphasizing the importance of source similarity over no similarity, different types of source similarity should be explored and compared. More studies should be conducted to identify the minimum effective number of personal characteristics necessary. Although no significant mediation effect was detected in this study, future studies could adopt relevant variables in its design to better detect the persuasive mechanism. No actual exercise data were collected in this study,

which constitutes a major limitation. Future studies could incorporate objective behavior measures collected from body sensors and accelerometers. The present research should also be replicated across different media platforms among different populations. Research has shown that people also prefer attending to arguments that highlight abstract rather than concrete features when attitude objects are temporarily distant [54]. In this study, almost all source similarity aspects were concrete and only a health behavior in the near future was examined. More research is needed to explore the feasibility of creating source similarity using abstract, or higher-level, personal characteristics (eg, personality characteristics such as introversion or extroversion) for different behaviors and other behavioral outcomes. Finally, the growing recognition of culture

as an important factor in health communication has the potential to contribute to the development of new and more effective message design. Although cultural values (eg, collectivism vs individualism) may not be inherently health-related, they may still influence health outcomes of the individuals and may enhance receptivity, acceptance, and salience of health messages [55].

To conclude, 2 types of messages (narrative and nonnarrative) and 3 types of source similarity (no similarity, non-health-related similarity, and health-related similarity) were empirically explored in a blog promoting the virtues of running. The results suggest that health-related source similarity is key for persuasive health communication especially when the messages are nonnarrative.

Conflicts of Interest

None declared.

References

1. Envision Solutions. Envision Solutions. 2008. Envision Solutions Insight Report: The evolving health blogsphere URL: http://www.envisionsolutionsnow.com/pdf/Studies/Health_Blogosphere_Final_sm.pdf [accessed 2008-10-10] [WebCite Cache ID 6HBRRUW3f]

2. Huang LS, Chou YJ, Lin CH. The influence of reading motives on the responses after reading blogs. Cyberpsychol Behav 2008 Jun;11(3):351-355. [doi: 10.1089/cpb.2007.0063] [Medline: 18537507]

3. Bialobzeskyte A. FounderTips. 2012. How Steve Kamb turned Nerd Fitness into a six-figure business URL: http://foundertips.com/blogging/how-steve-kamb-turned-nerd-fitness-into-a-six-figure-business [accessed 2012-12-12] [WebCite Cache ID 6AnurPtV9]

4. Rosenberg M. Conceiving the Self. New York: Basic Books; 1979.

5. Hinyard LJ, Kreuter MW. Using narrative communication as a tool for health behavior change: A conceptual, theoretical, and empirical overview. Health Educ Behav 2007 Oct;34(5):777-792. [doi: 10.1177/1090198106291963] [Medline: 17200094]

6. Green MC, Brock TC. In the mind's eye: Transportation-imagery model of narrative persuasion. In: Green MC, Strange JJ, Brock TC, editors. Narrative Impact: Social and Cognitive Foundations. Mahwah, NJ: Lawrence Erlbaum Associates; 2002:315-341.

7. Slater MD. Involvement as goal-directed strategic processing: Extending the elaboration likelihood model. In: Dillard JP, Pfau M, editors. The Persuasion Handbook: Developments in Theory and Practice. Thousand Oaks, CA: Sage Publications; 2002:175-194.

8. Eddens KS, Kreuter MW, Morgan JC, Beatty KE, Jasim SA, Garibay L, et al. Disparities by race and ethnicity in cancer survivor stories available on the web. J Med Internet Res 2009;11(4):e50 [FREE Full text] [doi: 10.2196/jmir.1163] [Medline: 19945948]

9. Chou WY, Hunt Y, Folkers A, Augustson E. Cancer survivorship in the age of YouTube and social media: a narrative analysis. J Med Internet Res 2011;13(1):e7 [FREE Full text] [doi: 10.2196/jmir.1569] [Medline: 21247864]

10. Meier A, Lyons EJ, Frydman G, Forlenza M, Rimer BK. How cancer survivors provide support on cancer-related Internet mailing lists. J Med Internet Res 2007;9(2):e12 [FREE Full text] [doi: 10.2196/jmir.9.2.e12] [Medline: 17513283]

11. McQueen A, Kreuter MW, Kalesan B, Alcaraz KI. Understanding narrative effects: the impact of breast cancer survivor stories on message processing, attitudes, and beliefs among African American women. Health Psychol 2011 Nov;30(6):674-682 [FREE Full text] [doi: 10.1037/a0025395] [Medline: 21895370]

XSL•FO

RenderX

12. Green MC, Brock TC. The role of transportation in the persuasiveness of public narratives. J Pers Soc Psychol 2000 Nov;79(5):701-721. [Medline: 11079236]

13. Rogers EM, Bhowmik DK. Homophily-heterophily: Relational concepts for communication research. Public Opin Q 1970;34(4):523-538. [doi: 10.1086/267838]

14. Andsager JL. Perceived Similarity of Exemplar Traits and Behavior: Effects on Message Evaluation. Communication Research 2006 Feb 01;33(1):3-18. [doi: 10.1177/0093650205283099]

15. Maccoby EE, Wilson WC. Identification and observational learning from films. J Abnorm Psychol 1957 Jul;55(1):76-87. [Medline: 13462664]

16. Bailenson JN, Yee N. Digital chameleons: automatic assimilation of nonverbal gestures in immersive virtual environments. Psychol Sci 2005 Oct;16(10):814-819. [doi: 10.1111/j.1467-9280.2005.01619.x] [Medline: 16181445]

17. Bott H. Observation of play activities in a nursery school. Genet Psychol Monogr 1928;4(1):44-88.

18. Brosius HB. Research note: The influence of exemplars on recipients' judgements: the part played by similarity between exemplar and recipient. European Journal of Communication 1999 Jun 01;14(2):213-224. [doi: 10.1177/0267323199014002004]

19. Simons HW, Berkowitz NN, Moyer RJ. Similarity, credibility, and attitude change: A review and a theory. Psychological Bulletin 1970;73(1):1-16. [doi: 10.1037/h0028429]

20. Taylor SE, Thompson SC. Stalking the elusive "vividness" effect. Psychological Review 1982;89(2):155-181. [doi: 10.1037/0033-295X.89.2.155]

21. Allen M, Preiss RW. Comparing the persuasiveness of narrative and statistical evidence using meta‐analysis. Communication Research Reports 1997 Mar;14(2):125-131. [doi: 10.1080/08824099709388654]

22. Petty RE, Cacioppo JT. The elaboration likelihood model of persuasion. In: Berkowitz L, editor. Advances in Experimental Social Psychology. New York: Academic Press; 1986:123-205.

23. Chaiken S. Heuristic versus systematic information processing and the use of source versus message cues in persuasion. Journal of Personality and Social Psychology 1980;39(5):752-766. [doi: 10.1037/0022-3514.39.5.752]

24. Slater MD, Rouner D. Entertainment-education and elaboration likelihood: Understanding the processing of narrative persuasion. Commun Theory 2002 May;12(2):173-191. [doi: 10.1111/j.1468-2885.2002.tb00265.x]

25. Maibach E, Flora JA. Symbolic modeling and cognitive rehearsal: using video to promote AIDS prevention self-efficacy. Communication Research 1993 Aug 01;20(4):517-545. [doi: 10.1177/009365093020004002]

26. Moyer-Gusé E. Toward a theory of entertainment persuasion: explaining the persuasive effects of entertainment-education messages. Communication Theory 2008;18(3):407-425. [doi: 10.1111/j.1468-2885.2008.00328.x]

27. Niederdeppe J, Shapiro MA, Porticella N. Attributions of responsibility for obesity: Narrative communication reduces reactive counterarguing among liberals. Human Communication Research 2011;37(3):295-323. [doi: 10.1111/j.1468-2958.2011.01409.x]

28. Green MC. Transportation theory. In: Donsbach W, editor. The International Encyclopedia of Communication. Malden, MA: Blackwell Publishing; 2008:5170-5175.

29. O'Keefe DJ. Persuasion: Theory and Research. Thousand Oaks, CA: Sage Publications; 2002.

30. Goodman N. Seven strictures on similarity. In: Goodman N, editor. Problems and Projects. Indianapolis, IN: Bobbs-Merrill; 1972:437-447.

31. Berscheid E. Opinion change and communicator-communicatee similarity and dissimilarity. Journal of Personality and Social Psychology 1966;4(6):670-680. [doi: 10.1037/h0021193]

32. Swartz TA. Relationship between source expertise and source similarity in an advertising context. Journal of Advertising 1984;13(2):49-55.

33. Mackie DM, Wright C. Social influence in an intergroup context. In: Brown R, Gaertner SL, editors. Blackwell Handbook of Social Psychology: Intergroup Processes. Malden, MA: Blackwell Publishers; 2001:281-300.

34. Green MC, Garst J, Brock TC. The power of fiction: Determinants and boundaries. In: Shrum LJ, editor. The Psychology of Entertainment Media: Blurring the Lines Between Entertainment and Persuasion. Mahwah, NJ: Lawrence Erlbaum Associates; 2004:161-176.

35. University of Minnesota. Study find lack of sleep, excessive computer screen time, stress and more hurt college students' grades. URL: http://www.bhs.umn.edu/healthdata/results/PressRelease_BoyntonOct2008.pdf [accessed 2013-06-06] [WebCite Cache ID 6HBUMGcQz]

36. American College Health Association. Healthy Campus 2010: Making it Happen. Baltimore, MD: American College Health Association; 2002.

37. Thomas RK, Terhanian G, Bayer LR. Behavioral intention measurement: International findings. In: Proceedings of Sixth International Conference on Logic and Methodology. 2004 Presented at: RC33 6th International Conference on Social Science Methodology: Recent Developments and Applications in Social Research Methodology, 2004; August 16-20, 2004; Amsterdam, The Netherlands p. 396-397.

38. O'Keefe DJ. Message properties, mediating states, and manipulation checks: claims, evidence, and data analysis in experimental persuasive message effects research. Commun Theory 2003 Aug;13(3):251-274. [doi: 10.1111/j.1468-2885.2003.tb00292.x]

XSL•FO

RenderX

39.   Kalyanaraman S, Sundar SS. The psychological appeal of personalized content in web portals: Does customization affect attitudes and behavior? J Communication 2006 Mar;56(1):110-132. [doi: 10.1111/j.1460-2466.2006.00006.x]

40.   Kovach CR. Content analysis of reminiscences of elderly women. Res Nurs Health 1991 Aug;14(4):287-295. [Medline: 1891614]

41.   Cacioppo JT, Petty RE. Social psychological procedures for cognitive response assessment: The thought-listing technique. In: Merluzzi TV, Glass CR, Genest M, editors. Cognitive Assessment. New York: Guilford; 1981:309-342.

42.   Metzger MJ, Flanagin AJ, Eyal K, Lemus DR, McCann RM. Credibility for the 21st century: Integrating perspectives on source, message, and media credibility in the contemporary media environment. Communication Yearbook 2003;27:293-335.

43.   McCroskey JC, McCain TA. The measurement of interpersonal attraction. Speech Monographs 1974 Aug;41(3):261-266. [doi: 10.1080/03637757409375845]

44.   Eyal K, Rubin AM. Viewer aggression and homophily, identification, and parasocial relationships with television characters. Journal of Broadcasting & Electronic Media 2003 Jan;47(1):77-98. [doi: 10.1207/s15506878jobem4701_5]

45.   Rubin AM. Parasocial interaction scale. In: Rubin RB, Palmgreen P, Sypher HE, editors. Communication Research Measures: A Sourcebook. New York: Guilford Press; 1994:273-277.

46.   Green MC. Transportation into narrative worlds: The role of prior knowledge and perceived realism. Discourse Processes 2004 Sep;38(2):247-266. [doi: 10.1207/s15326950dp3802_5]

47.   Meijnders A, Midden C, Olofsson A, Ohman S, Matthes J, Bondarenko O, et al. The role of similarity cues in the development of trust in sources of information about GM food. Risk Anal 2009 Aug;29(8):1116-1128. [doi: 10.1111/j.1539-6924.2009.01240.x] [Medline: 19473312]

48.   Jiang L, Hoegg J, Dahl D, Chattopadhyay A. The persuasive role of incidental similarity on attitudes and purchase intentions in a sales context. J CONSUM RES 2010 Feb;36(5):778-791. [doi: 10.1086/605364]

49.   Hilmert CJ, Kulik JA, Christenfeld NJ. Positive and negative opinion modeling: the influence of another's similarity and dissimilarity. J Pers Soc Psychol 2006 Mar;90(3):440-452. [doi: 10.1037/0022-3514.90.3.44] [Medline: 16594830]

50.   Ryan RM, Deci EL. On happiness and human potentials: a review of research on hedonic and eudaimonic well-being. Annu Rev Psychol 2001;52:141-166. [doi: 10.1146/annurev.psych.52.1.141] [Medline: 11148302]

51.   Brehm SS, Brehm JW. Psychological Reactance: A Theory of Freedom and Control. New York: Academic Press; 1981.

52.   Beniger JR. Personalization of mass media and the growth of pseudo-community. Communication Research 1987 Jun 01;14(3):352-371. [doi: 10.1177/009365087014003005]

53.   Gardner WL, Gabriel S, Lee AY. "I" value freedom, but "we" value relationships: Self-construal priming mirrors cultural differences in judgment. Psychological Science 1999 Jul 01;10(4):321-326. [doi: 10.1111/1467-9280.00162]

54.   Fujita K, Eyal T, Chaiken S, Trope Y, Liberman N. Influencing attitudes toward near and distant objects. J Exp Soc Psychol 2008;44(3):562-572. [doi: 10.1016/j.jesp.2007.10.005]

55.   Kreuter MW, McClure SM. The role of culture in health communication. Annu Rev Public Health 2004;25:439-455. [doi: 10.1146/annurev.publhealth.25.101802.123000] [Medline: 15015929]

Abbreviations

ANCOVA: analysis of covariance
ANOVA: analysis of variance
HSD: honestly significant difference

XSL•FO
RenderX

XSL•FO
RenderX

<u>Original Paper</u>

Interactive Algorithms for Teaching and Learning Acute Medicine in the Network of Medical Faculties MEFANET

Daniel Schwarz[1], MSc, PhD; Petr Štourač[1,2], MD, PhD; Martin Komenda[1], MSc, Dr; Hana Harazim[1,2], MD; Martina Kosinová[2], MD; Jakub Gregor[1], MSc, PhD; Richard Hůlek[1], BSc; Olga Smékalová[2], MD; Ivo Křikava[2], MD; Roman Štoudek[2], MD; Ladislav Dušek[1], MSc, Dr, PhD

[1]Institute of Biostatistics and Analyses, Faculty of Medicine, Masaryk University, Brno, Czech Republic

[2]Department of Anesthesiology and Intensive Care Medicine, University Hospital Brno, Faculty of Medicine, Masaryk University, Brno, Czech Republic

Corresponding Author:
Daniel Schwarz, MSc, PhD
Institute of Biostatistics and Analyses
Faculty of Medicine
Masaryk University
Kamenice 5
Brno, 625 00
Czech Republic
Phone: 420 549492854
Fax: 420 549492855
Email: schwarz@iba.muni.cz

Abstract

Background: Medical Faculties Network (MEFANET) has established itself as the authority for setting standards for medical educators in the Czech Republic and Slovakia, 2 independent countries with similar languages that once comprised a federation and that still retain the same curricular structure for medical education. One of the basic goals of the network is to advance medical teaching and learning with the use of modern information and communication technologies.

Objective: We present the education portal AKUTNE.CZ as an important part of the MEFANET's content. Our focus is primarily on simulation-based tools for teaching and learning acute medicine issues.

Methods: Three fundamental elements of the MEFANET e-publishing system are described: (1) medical disciplines linker, (2) authentication/authorization framework, and (3) multidimensional quality assessment. A new set of tools for technology-enhanced learning have been introduced recently: Sandbox (works in progress), WikiLectures (collaborative content authoring), Moodle-MEFANET (central learning management system), and Serious Games (virtual casuistics and interactive algorithms). The latest development in MEFANET is designed for indexing metadata about simulation-based learning objects, also known as electronic virtual patients or virtual clinical cases. The simulations assume the form of interactive algorithms for teaching and learning acute medicine. An anonymous questionnaire of 10 items was used to explore students' attitudes and interests in using the interactive algorithms as part of their medical or health care studies. Data collection was conducted over 10 days in February 2013.

Results: In total, 25 interactive algorithms in the Czech and English languages have been developed and published on the AKUTNE.CZ education portal to allow the users to test and improve their knowledge and skills in the field of acute medicine. In the feedback survey, 62 participants completed the online questionnaire (13.5%) from the total 460 addressed. Positive attitudes toward the interactive algorithms outnumbered negative trends.

Conclusions: The peer-reviewed algorithms were used for conducting problem-based learning sessions in general medicine (first aid, anesthesiology and pain management, emergency medicine) and in nursing (emergency medicine for midwives, obstetric analgesia, and anesthesia for midwifes). The feedback from the survey suggests that the students found the interactive algorithms as effective learning tools, facilitating enhanced knowledge in the field of acute medicine. The interactive algorithms, as a software platform, are open to academic use worldwide. The existing algorithms, in the form of simulation-based learning objects, can be incorporated into any educational website (subject to the approval of the authors).

XSL•FO
RenderX

Introduction

Medical education is constantly evolving by gradually, but significantly, shifting from traditional methods (eg, textbooks, lectures, bedside teaching) to a more comprehensive approach that also employs modern information and communication technology (ICT) tools (eg, e-learning, interactive algorithms, computer simulations, virtual patients). Such approaches have been demonstrated to enhance and improve the learning skills of medical students and residents in comparison to traditional methods [1-3]. Several ancillary factors in medicine and medical education have also contributed significantly to these trends; in particular, the rapid development of new technologies and the generally preferred shorter hospital stays, which reduces the student's exposure to a given case or diagnosis. The economic efficiencies of Web-based education and traditional face-to-face education approaches were compared under randomized controlled trial conditions in Maloney et al [4] and it was shown that the Web-based education approach was clearly more efficient from the perspective of the education provider.

Although most of the modern interactive tools are intended for extending and supplementing the traditional methods rather than replacing them, they have undoubtedly brought a number of advantages, such as equal and easy access for the students to all diagnoses, simulation of a variety of real-life situations, comprehensive interdisciplinary learning, and a higher level of comfort for hospitalized patients. Simulation-based learning also provides the unique opportunity of practicing knowledge application in a manner that mimics real-time patient care without posing a risk to the patient [5,6]. On the other hand, developing simulations and e-learning materials requires investment of the time of skilled professionals (eg, physicians, teachers, programmers); therefore, it is necessary to ensure that the time and resources expended is justified by the educational impact [7]. Furthermore, the developed tools are often accepted uncritically and with emphasis on technological sophistication at the expense of the underlying psychopedagogical theories [1].

Improved efficiency in the development of digital teaching and learning materials, as well as their higher quality, can be achieved by sharing the educational content and by initiating collaborative multi-institutional authoring teams together with joint efforts in establishing the methods for quality evaluation. The management of multisource content among academic institutions brings the necessity of correct indexing, metadescription, and proper categorization [8], as well as reimbursement [9-11] for the created resources. The idea of the medical faculties in the Czech Republic and Slovakia sharing their educational digital contents surfaced in 2006 for the first time. Soon after, in 2007, all 7 Czech medical faculties as well as all 3 Slovak medical faculties formally joined the new network. In 2012, representatives of the Czech and Slovak health care institutions joined the Medical Faculties Network (MEFANET) education network. The MEFANET project [12] aims to develop cooperation among the medical faculties to further the education of medical and health care disciplines using modern ICT via a common platform for sharing digital education content, as well as for assessing their quality through a multidimensional approach [13].

Most of the digital teaching described in recent literature has been prepared as Web-based works because Web technologies allow for easy incorporation of multimedia objects, interactive algorithms, animated simulations, etc. The work may then be easily accessed from any computer and by a defined target audience (eg, students of a particular medical school or course). The developed tools and simulations cover a wide range of medical disciplines, such as critical care [14,15], cardiology [3], hematology [1], neurology [16], surgery [17], metabolic disorders, imaging methods [18,19], and cytogenetics [20].

Acute medicine is a dynamic environment with high demands on team communication and leadership, requiring correct clinical reasoning and quick decision making under time pressure. Simulation offers a good and interesting platform for training multidisciplinary medical teams, facilitating interaction among the team members and enabling the team to function in an effective and coordinated manner [6]. Internet education resources for intensive care medicine have recently been reviewed by Kleinpell et al [14], who demonstrated that most of them are electronic forms of textbooks and articles rather than interactive algorithms and dynamic simulations. Davids et al [7] described an interactive Web-based simulation in which the user treats patients with electrolyte and acid-base disorders, selects the therapies and doses, and can immediately see the treatment results.

In this paper, we present the education portal AKUTNE.CZ [21] as an important part of the MEFANET's contents. It aims to be a comprehensive source of information and education materials covering all aspects of acute medicine for undergraduate and postgraduate students of the medical and health professions. We focus here primarily on the simulation-based tools for teaching and learning algorithms for acute patient care that form the backbone of AKUTNE.CZ. The simulations take the form of interactive algorithms and represent the basis for a new extension of MEFANET's activities incorporating focus on serious games.

Methods

Overview

MEFANET [12] has established itself as the standard-setting body for medical educators in the Czech Republic and Slovakia, 2 independent countries that once comprised a federation, have similar languages, and still retain the same curricular structure for medical education. One of the basic goals of the network is to advance medical teaching and learning with the use of modern information and communication technologies. As an instrument, MEFANET has decided to develop an original and uniform solution for educational Web portals that are used, together with a central gateway, to offer and share digital education content.

Students—approximately 16,500 potential users and academic staff and approximately 3900 potential users from all Czech and Slovak medical faculties—can find their e-learning materials at 11 standalone faculties' instances of an educational portal with the use of the indexing and searching engine, MEFANET Central Gateway [22].

MEFANET e-Publishing System

The idea of a shared e-publishing system is based on a set of standalone Web portals rather than on a centralized application hosted for all medical schools, which might be an inflexible and more vulnerable alternative solution. Each portal instance represents an independent publication media with its own International Standard Serial Number (ISSN) code and an editorial board. Local metadata describing the digital educational contents are replicated regularly to the central gateway (see metadata harvesting in Figure 1). There are 3 fundamental elements that have to be rigidly maintained on the part of local administrators: (1) the medical disciplines linker, (2) the authentication/authorization framework, and (3) multidimensional quality assessment. The other features, properties, and functionalities can be adapted or localized to meet the needs of the particular institution. A detailed description of the 3 fundamental elements is as follows. See [13] for full and comprehensive information.

The medical disciplines linker represents the main taxonomy of contributions within the frame of the network. With its single-level list of 56 medical specializations, it forms the only obligatory structure of a portal instance. Any change to its content is subject to approval of the MEFANET Coordinating Committee.

The authors of the shared teaching materials can choose from the following user groups to permit or deny access to their materials: (1) nonregistered anonymous users, (2) registered anonymous users who accept the terms of use within their registration, (3) users of the MEFANET network, that is, a student or teacher from any Czech or Slovak medical school (MEFAPERSON), (4) users from a local university whose affiliation to that university has been verified at the portal via the local information system of that university, (5) users to whom attachments are made available only after the author's explicit consent. Services of the Czech academic identity federation, eduID.cz [23], are used to check the affiliations of the users of the portal instances. This federation uses the Shibboleth technology, which is one of the several authentication frameworks allowing the sharing of Web resources among institutions using the Security Assertion Markup Language (SAML) protocol standard. The portal instances behave like

service providers in this federation, whereas the information systems of the involved schools act as identity providers.

There are 4 dimensions of critical importance when evaluating the quality of electronic teaching materials: (1) expert review, (2) education level of target users, (3) classification by type, and (4) self-study score. The review includes binary questions as well as open questions. The structure of the review form can be localized by modifying an extensible markup language (XML) template file. The second dimension is represented by the education level of the target group of the teaching material, which is a useful piece of information for the users and the reviewers. The next dimension is represented by a multiple-choice classification according to the types of attachments—the enumerated scale includes static files for Web-based learning and interactive e-learning courses encapsulated in the learning management systems. The last dimension—a self-study score—indicates what users think about the usability of a particular contribution in their self-studies. The values of the first 3 dimensions of the 4D assessment are composed by authors, guarantors, and reviewers. Their activities and the workflow of a contribution are explained in Figure 2. In addition to the 4D quality assessment, all contributions submitted to the central gateway undergo an additional editorial process called *mentally active monitoring*. It focuses on the following issues: (1) metadata is filled in properly, (2) granularity of the attachments is suitable, and (3) all attached documents and the links are accessible for at least MEFAPERSON users. The monitoring of these 3 important issues is done not only at the syntax level, but also semantically; therefore, it is carried out by a team of editors in cooperation with the editors responsible for the local Web portals.

Recently, new tools for technology-enhanced learning have been introduced to the MEFANET network in addition to the common e-publishing portal platform. These new tools complement the portal platform suitably because they provide a higher level of interactivity for students during their self-study process. Figure 1 shows how the new 4 tools—Sandbox [24], WikiLectures [25], Moodle-MEFANET [26], and Serious Games [27]—are related to the already established and standardized MEFANET Central Gateway.

The Serious Games extension is the latest development in MEFANET and it is designed for indexing metadata about simulation-based learning objects, also known as *electronic virtual patients* or *virtual clinical cases*. The first comprehensive set of such interactive learning objects is composed by algorithms for acute patient care published at the AKUTNE.CZ educational portal [21] together with other digital education materials covering a wide range of acute medicine topics.

Figure 1. MEFANET involves all medical schools in the Czech Republic and Slovakia. They share one another's digital teaching and learning materials by using an e-publishing system that consists of 11 educational Web portals and a central gateway. The extensions of the MEFANET e-publishing system appear as standalone platforms for their users. However, all teaching or learning materials indexed by the MEFANET Central Gateway undergo the same procedures of multidimensional quality assessment.
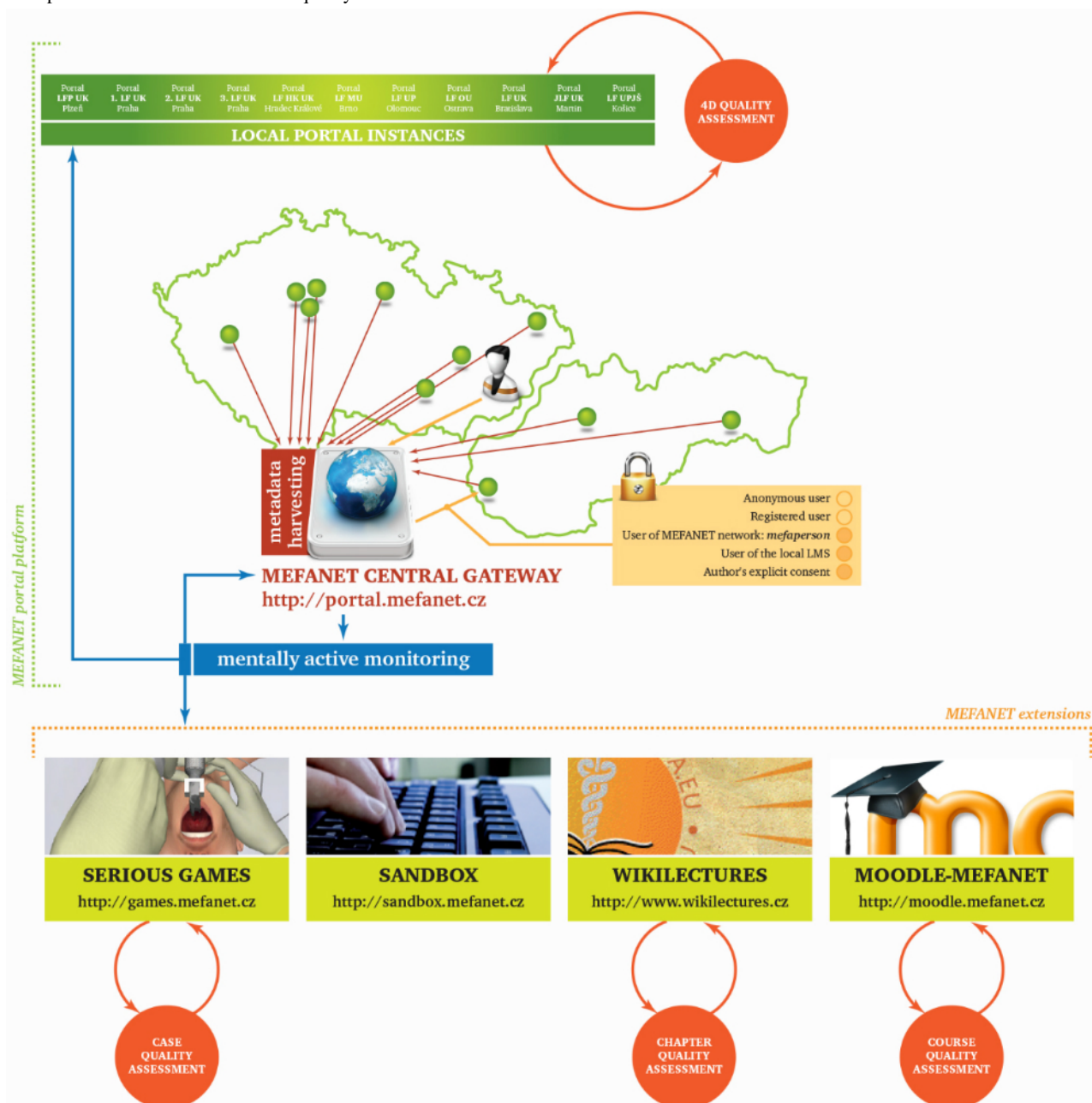
Figure 2. The contribution workflow scheme: (1) the author and technical editor finishes the contribution, (2) the guarantor, who is associated with a particular medical discipline, is notified about a new contribution to his/her field of interest, (3) the guarantor, either alone or with the help of the faculty's editorial committee, invites 2 reviewers to present their reviews online with the use of template-generated forms.
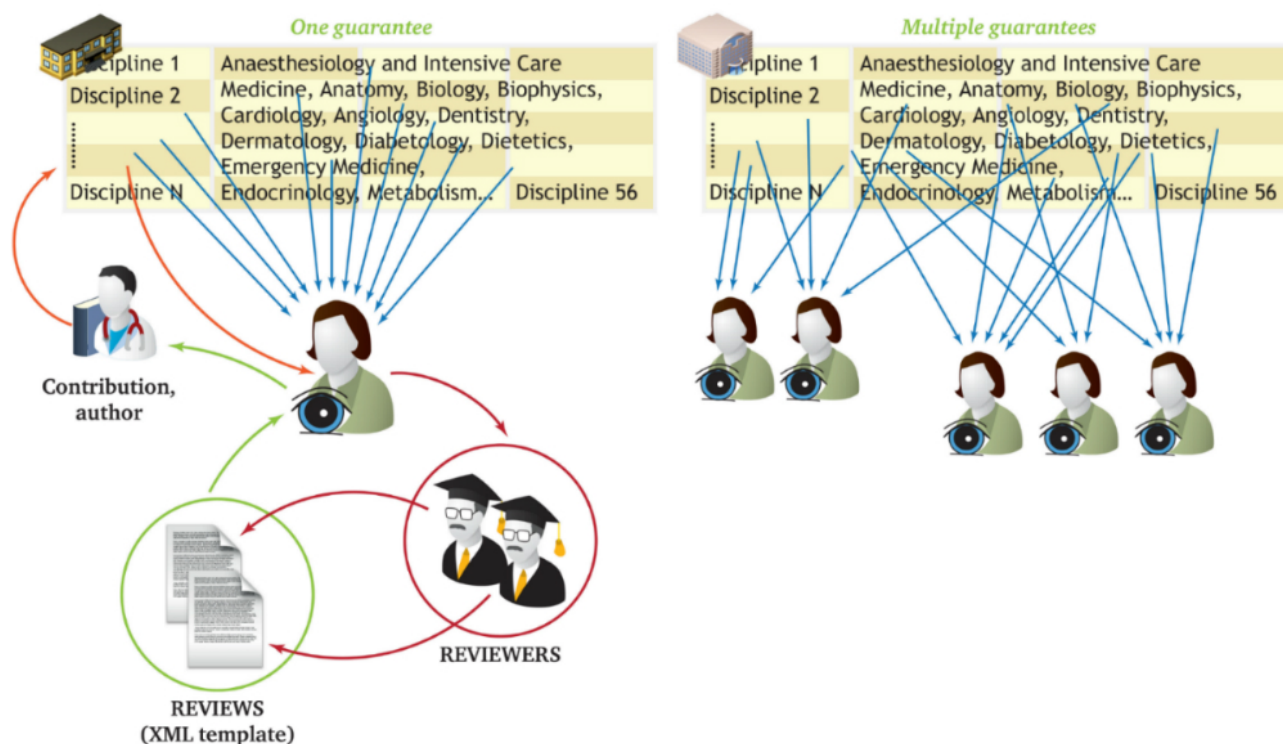
Interactive Algorithms for Teaching and Learning Acute Medicine

Each physician dealing with acute patients needs algorithmic thinking and correct clinical reasoning. Our interactive algorithms take the form of content-rich virtual cases because they link together process flowcharts and multimedia. Creating such algorithms or electronic virtual patients is laborious, time-consuming, and often accompanied by ambiguities and hesitations. Following the principles of student-centered learning, our authoring teams consisted of medical students in the final years of their studies, supervised by an experienced clinician. The complete workflow of the authoring process is outlined in Figure 3.

It takes 10 to 50 hours of active work to produce 1 interactive algorithm. The time of the team members is spent on collaborative work, meetings, and on self-studying. Student–authors consult their problems and reservations with a supervisor assigned to them and the resulting product is then submitted to an external reviewer, usually an experienced clinician or an academic staff from another workplace. After the incorporation of all reviewers' comments, the algorithm is completed by metadata to be published on the AKUTNE.CZ educational portal. Finally, sets of algorithms are compiled together with their metadata into a contribution to be published and indexed on the MEFANET Central Gateway. These contributions with a wider scope than individual algorithms are subjected to the multidimensional quality assessment described previously. Finished and published algorithms are used by other students either as outlines for problem-based learning (PBL) sessions or as supplementary materials for training and adopting correct clinical reasoning.

The interactive algorithms are authored with the use of a Web-based (PHP/MySQL) BackOffice application that provides the student–authors the following functionalities through its online forms and drag and drop control: (1) node-based scenario design, (2) description of the situation in each node, including the intervals of parameter values of physical examinations, intervals of laboratory values, and multimedia, (3) description of the correct answers as well as distractors with the option to repeat or end in a fatality, and (4) data export for each finished algorithm into an XML document. The XML documents are then rendered into a Flash object resembling a serious game. A student–player uses the game or this simulation-based learning object by moving between the nodes, which may be of different types, as shown in the sample algorithm in Figure 4. Each move causes a shift in the timeline as a side effect of the student–player's action, lending authenticity to the scenario and creating a stress effect, which is pronounced in real-life situations when dealing with acute patients. Continuous change of various numerical parameters reflecting the development of patient's clinical status and vital functions in time (eg, blood pressure, pulse, oxygen saturation) is also available (see the example of a node of a selected algorithm in Figure 5).

Students' Feedback on the Interactive Algorithms

We asked students about their attitudes and interest in using the interactive algorithms as part of their medical or health care studies. The purpose was to ascertain how the students perceived our efforts on authoring and implementing simulation-based learning tools that are so demanding to create. An anonymous questionnaire of 10 items (see Table 1 for complete overview of questions and answer options) was created and presented via SurveyMonkey [28], a free online survey software. Data

collection lasted for 10 days in February 2013. The students who enrolled at 1 of the educational workshops or a conference organized by the group around the AKUTNE.CZ portal were asked to complete the survey. The first 4 questions were aimed at obtaining basic data about the respondents, so that the ones who did not study any field of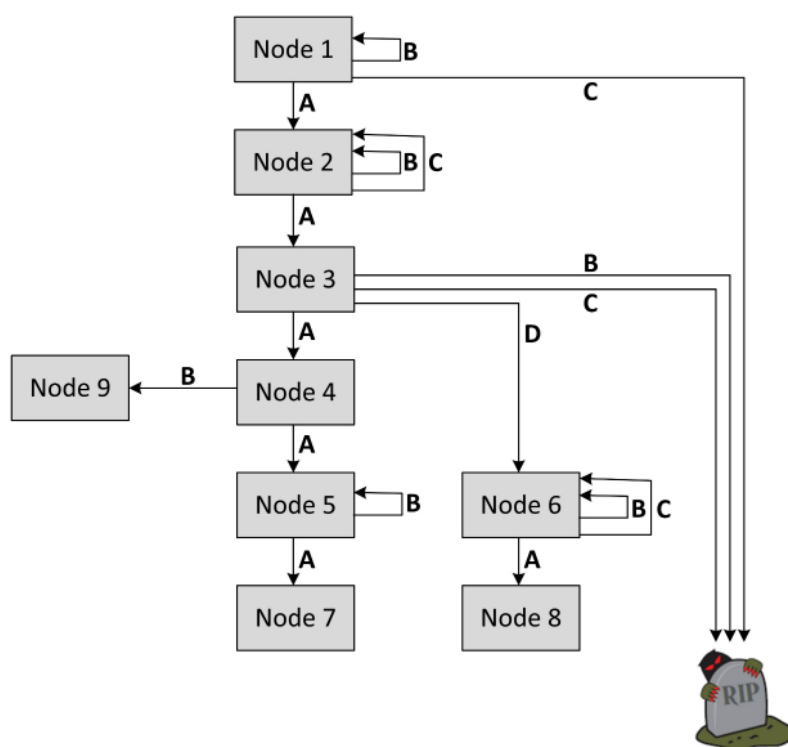 medicine or health care could be filtered out as well as the ones who did know about our interactive algorithms at all. Further questions were answered with a 5-point Likert scale and 1 binary question was aimed at seeking feedback on the use of our interactive algorithms in the studies of acute medicine topics.

Figure 3. The authoring workflow of an interactive algorithm from choosing the topic through a review process to deployment to teaching in the form of a moderated problem-based learning session.
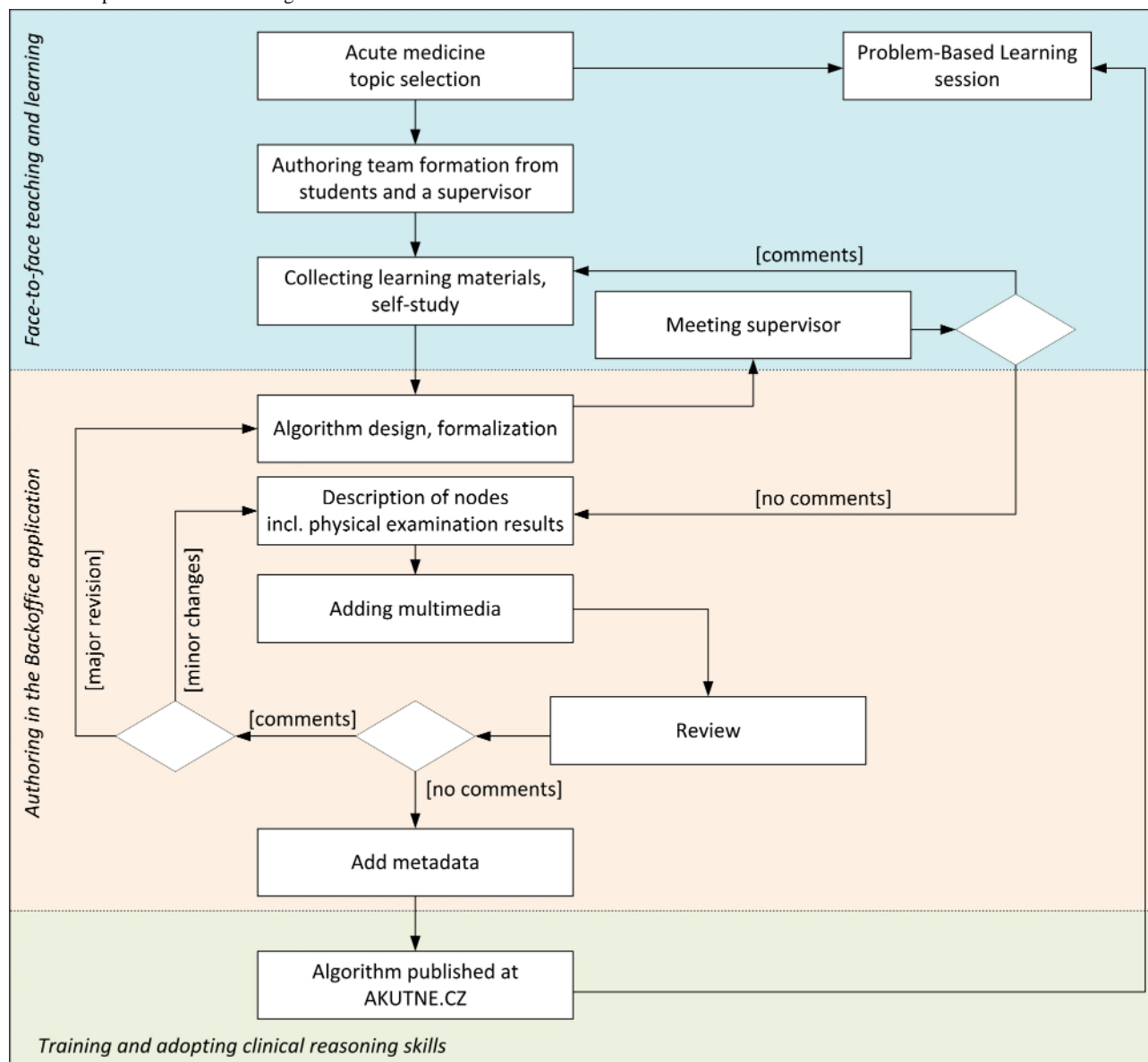
Figure 4. Various types of nodes and options/answers that may be used for authoring an interactive algorithm.

Figure 5. An explained screenshot for 1 node of an algorithm for training clinical reasoning skills in acute coronary syndrome.

XSL•FO
RenderX

Table 1. Questionnaire for collecting the students' feedback on the interactive algorithms.

#	Question	Answer options
1	State your gender.	Male or female
2	What is your field of study?	General medicine
		Dentistry
		Health care specializations (MSc)
		Health care specializations (BSc)
		Midwifery (BSc)
		Postgraduate doctoral program. another (specify, please)
3	What is your attitude toward the interactive algorithms AKUTNE.CZ?	I do not know what they are
		I know what they are, but I have never used them
		I tried to solve at least 1 interactive algorithm
		I am an author or a coauthor of at least 1 interactive algorithm
4	Have you ever used for your studies a serious game (simulation of real situations for teaching and learning) or-any other interactive algorithm AKUTNE.CZ?	I have not used any at all, not even any interactive algorithm
		No. I have used only the interactive algorithms
		Yes. I have used also...(specify which):
5	The interactive algorithms AKUTNE.CZ are an effective tool for my learning.	5-point Likert scale from strongly disagree to strongly agree
6	The use of the interactive algorithms AKUTNE.CZ improved my knowledge in the field of acute medicine.	5-point Likert scale from strongly disagree to strongly agree
7	The use of the interactive algorithms AKUTNE.CZ represents for me a better way to study than static textbooks.	5-point Likert scale from strongly disagree to strongly agree
8	I like playing the interactive algorithms AKUTNE.CZ not only at home, but also at school under the supervision of teachers, together with consulting possible answers as well as with discussion on all issues related to the topic.	5-point Likert scale from strongly disagree to strongly agree
9	Multimedia accompanying the decision nodes together with the time stressor evokes an authentic atmosphere of clinical reasoning and decision making.	5-point Likert scale from strongly disagree to strongly agree
10	Would you recommend the interactive algorithms AKUTNE.CZ to your friends?	Yes or no

Results

Over 5 years, almost 25 interactive algorithms in the Czech and English languages have been developed and published on the AKUTNE.CZ educational portal to allow the users to test and improve their knowledge and skills in the field of acute medicine. Another 5 algorithms will be finished during 2013. They cover a wide range of acute medicine topics in the following 5 packages:

Basic Life Support and Advanced Life Support

Algorithms cover many basic life support (BLS) and advanced life support (ALS) procedures described in the current European Resuscitation Council guidelines. We developed a BLS for adults algorithm, ALS for bradycardia, BLS for choking children, and a foreign-body airway obstruction in adults algorithm.

Emergency Medicine

Emergency medicine is a very specific type of care in exceptional conditions. We tried to create an ambience of a real car accident in the interactive algorithm. Further topics of emergency medicine are algorithms for water rescue, severe hypothermia in the mountains in winter, out-of-hospital craniocerebral injury, and syncope.

Critical Care Medicine

Critical care medicine (CCM) is the flagship of medicine in general. It is no coincidence that the most demanding and complex algorithms are from this field. The surviving sepsis algorithm is based on the surviving sepsis guidelines of the Society of Critical Care Medicine (SCCM). The acute coronary syndrome algorithm provides a complete decision tree for a patient with acute myocardial stroke. The algorithm for diabetes mellitus deals with sudden loss of consciousness in a diabetic patient.

XSL•FO
RenderX

Anesthesiology

These algorithms cover both interesting acute and propaedeutic situations during anesthesia. We developed an algorithm describing the correct approach to the parturient with postdural puncture headache after epidural labor analgesia. Another acute situation is described in the algorithm for toxic reaction to anesthetic agents. Propaedeutic skills are represented by algorithms introducing the insertion of central venous catheter or the choosing of venous entry routes.

Pain Management

Providing good analgesia for acute and chronic pain is a global issue. We cover these issues with an acute postoperative pain algorithm and by algorithms with correct approach to analgesia in a general practitioner's and a dentist's surgery/clinic.

User's attendance to the interactive algorithms was analyzed with the use of Google Analytics in context of the whole website AKUTNE.CZ within a 1-month period (January 15 to February 14, 2013). In this period, 3342 unique users visited the website (5452 visits in total, 176 visits per day, SD 53.1). All interactive algorithms together had 816 unique users. Of 816 users, 297 (36.4%) accessed the algorithms from Brno and were, therefore, identified as students of the Faculty of Medicine in Brno. Other large groups of visitors were from Prague (99/816, 12.1%) and Bratislava (26/816, 3.2%), both major cities with established medical education facilities. On the other hand, 259 accesses (31.7%) were from places where no faculty of medicine exists. Although we are aware of the limited information value of such analysis (eg, not all visits from Brno are performed at school, or a visitor from a small village could be a student from the Brno faculty of medicine), these results document that the interactive algorithms have been used within the whole MEFANET network and a significant proportion of students use them in places outside of the school (ie, in their homes and during leisure time). The most frequently played algorithms

were the diabetes mellitus (94/816 unique users, 11.5%), hypothermia (89 unique users, 10.9%), and surviving sepsis (52 unique users, 6.4%).

In the feedback survey, 62 participants (13.5%) completed the online questionnaire out of the overall 460 asked to participate. Of all respondents, 66.1% were women and 33.9% were men. After filtering out the participants who were not students of any medical or health care program, and those who did not know about the availability of the interactive algorithms AKUTNE.CZ, the resulting responses from 54 participants were analyzed (see Figure 6). The participants were asked whether the interactive algorithms served as an effective tool for their learning. Four responses were negative or very negative (7.4%), 3 responses were neutral (5.6%), and 47 responses were positive or very positive (87.0%). The participants were further asked whether the interactive algorithms improved their knowledge of acute medicine. Six responses were negative or very negative (11.2%), 4 responses were neutral (7.4%), and 44 responses were positive or very positive (81.4%). In all, 40 participants agreed or strongly agreed (74.0%) that the interactive algorithms represented for them a better study method in comparison to static textbooks, whereas 6 participants disagreed or strongly disagreed (11.2%), and a further 8 respondents neither agreed nor disagreed (14.8%). The participants' attitude toward interactive algorithms as a tool for face-to-face teaching and learning was positive or very positive in 46 responses (85.2%), negative or very negative in 3 responses (5.6%), and neutral in 5 responses (9.2%). Most participants agreed or strongly agreed (47/54, 87.0%) that multimedia and the time-stress factor provided an authentic atmosphere for pertinent clinical reasoning, whereas 4 participants disagreed or strongly disagreed (7.4%) with this fact and 3 were unsure (5.6%). All participants (100%) stated that the interactive algorithms were worth recommending to their friends.

Figure 6. Attitudes and interests of students about using the interactive algorithms as part of their medical or health care studies.
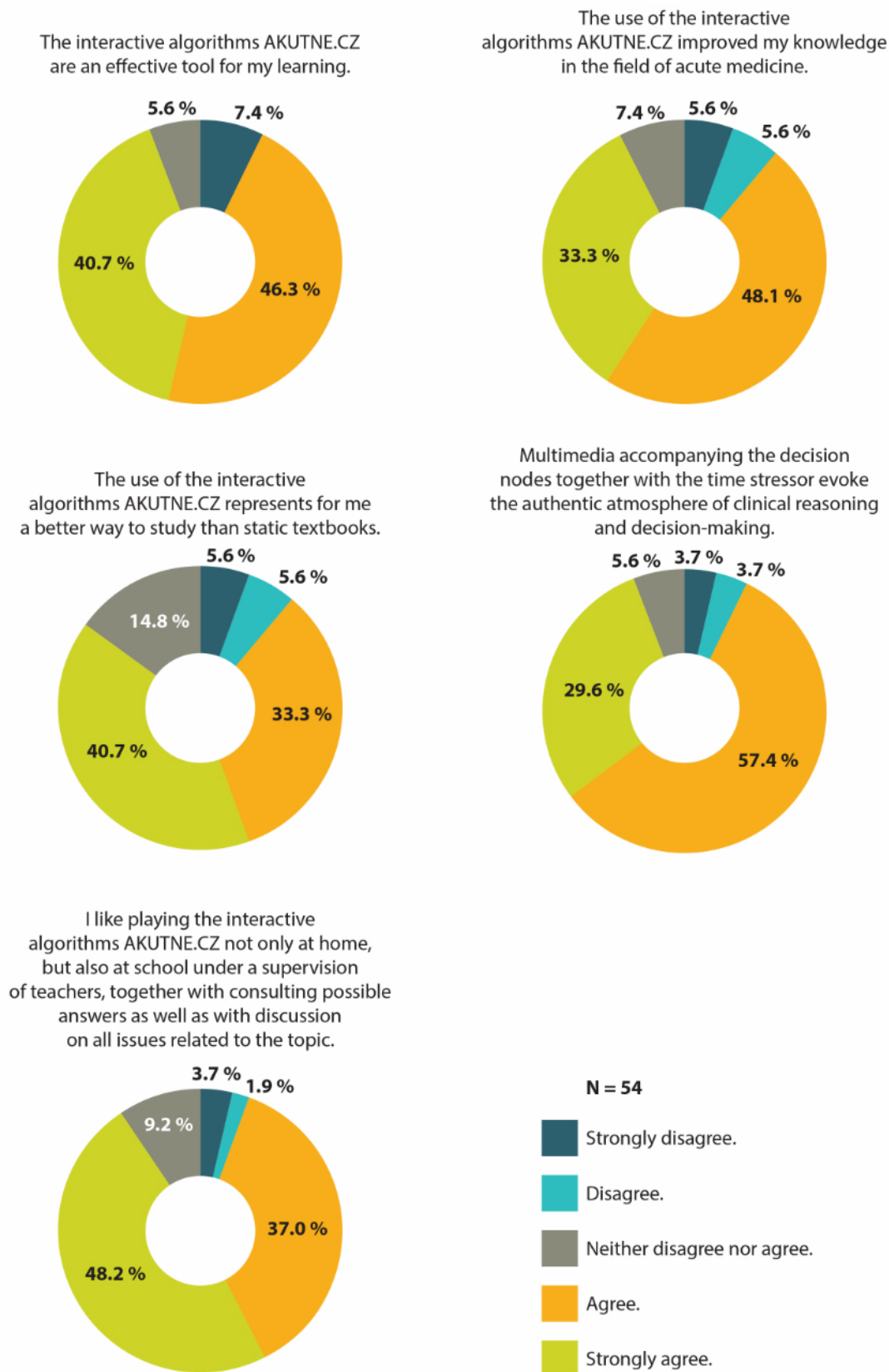
Discussion

Principal Findings

High-quality digital education content production has become a matter of prestige at medical schools in the Czech Republic and Slovakia, and the volume of teaching and learning materials available is growing rapidly thanks to the MEFANET project and its ICT platforms, which have been continuously developed and adopted to the needs of the MEFANET community during the past 6 years. Four new extensions, which complement the e-publishing portal platform standardized in MEFANET, are usable independently; however, their complex application in conjunction with the portal platform as a tool for final e-publishing will allow more effective repurposing of the materials created with the use of the extensions, as well as broader integration of the digital education contents among the MEFANET community. Further development aims to encourage the publication of materials for the teaching of clinical reasoning based on the concept of interactive algorithms or virtual patients. Such simulation-based learning objects are aimed to help the student in developing the much-needed confidence to manage acute conditions, to react accurately, and to avoid distraction by secondary issues.

The unique advantage of interactive algorithms AKUTNE.CZ is the possibility to create complex and branching scenarios. Nevertheless, real-life medical emergencies offer little or no extra options; in many cases, there is only 1 correct course of action. Unfortunately, this feature has not been adopted on a wide scale. The reason could be the characteristics of real-time acute medicine situations that are often linear with no space for branching. On our part, we have complied as much as possible with the guidelines of medical societies. Any deviation from the approved procedures may lead to deteriorating outcomes in real clinical situations. This is the reason why we prefer creating simplified and linear algorithms. An algorithm that approaches realistic simulation (nonlinear or open format) could be more attractive for the students, but we believe that to happen at the expense of didacticism. We also prefer topics that are endorsed and processed by the guidelines or recommendations of the European medical societies (ie, European Resuscitation Council, SCCM, European Society of Regional Anaesthesia and Pain Therapy) and/or national medical societies (ie, Czech Society of Anaesthesiology and Intensive Care Medicine, Czech Society of Intensive Care Medicine, Czech Society of Hematology, Czech Society of Cardiology, Czech Gynecological and Obstetrical Society, and Czech Pain Society). The linear scenarios help to maintain a didactic focus of the interactive algorithms. This mechanistic approach may, however, be detrimental to the students' understanding of the underlying physiological processes. In order to overcome this limitation, we prefer to use the interactive algorithms for teaching in the form of moderated PBL sessions. Inspired by several works in the field of advanced physiological simulators with a mathematical background [29-31], we will focus our future developments toward a technology mashup, which would allow to incorporate time-dependent, complex physiological simulation of multiple variables and their response to perturbations into the multimedia part of the interactive algorithms.

We cover a wide range of acute medicine topics through the AKUTNE.CZ algorithms. Of course, there is room for additional themes, for example, the widely publicized case of methanol poisoning in 2012 in the Czech Republic, which led to fatalities. Other topics under consideration include selected amyotrophic lateral sclerosis scenarios and out-of-hospital medical emergencies. Interactive algorithms are also used during obstetric anesthesia and analgesia lessons for the midwives—severe peripartal bleeding, amniotic fluid embolism, and out-of-hospital delivery algorithms. The primary aim is to achieve a situation whereby each acute medicine teaching unit has at least 1 interactive algorithm for PBL.

Although the algorithms were tailored to the teaching and learning of acute medicine issues, it is possible to use them for education in other medical and health care disciplines as well. The selection of the parameters from physical examination results and laboratory tests can be changed easily and, thus, adopting the tool for use elsewhere. In comparison with other examples of simulation-based learning objects, such as virtual patients [32], we have a different approach to handling the selected physical examination results and laboratory findings. We follow real-world scenarios and provide the possibility to record these parameters as they are recorded during management of real acute patients too. Each measurement is linked to an increase of the time-stress factor. Thus, students not only learn about dynamics of these characteristics, but also about the unpleasant price in terms of time spent for unnecessary measurements.

A major problem with any medical issue is topicality. AKUTNE.CZ algorithms overcome such problems by ensuring regular updates through the combined efforts of medical students and the authors, in addition to holding regular meetings on time-scheduled updated topics. The algorithms truly reflect on the current medical recommendations and guidelines of the medical societies.

In general, our survey points to a fairly strong preference for the AKUTNE.CZ interactive algorithms by the students as part of their medical or health care studies, although it is notable that the participants were only just aware of the interactive algorithms—a small proportion (9%) reported using other serious games or simulation-based learning objects for their studies. Nevertheless, positive attitudes toward the interactive algorithms outnumbered negative responses. Confirming our expectations, one of the strongest positive answers concerned the participants' desire to use the interactive algorithms not only for their self-studies during leisure time, but also in face-to-face teaching and learning. Based on our several preliminary attempts at implementing the PBL principles into our teaching, we are fully confident about PBL-like sessions conducted on the node-based scenarios of selected interactive algorithms as the appropriate way to fulfill that wish. The medical and health care institutions in the Czech Republic and Slovakia involved in MEFANET are currently, however, in the very preliminary phases of implementing PBL into their curriculum. Hopefully, the use of interactive algorithms in the process of PBL implementation shall pave the way toward increased attractiveness of our teaching, as well as deeper interest on the part of the students not only in acute medicine issues.

XSL•FO

RenderX

Limitations

A limitation of the study is that we did not collect data to observe effects of the use of algorithms on expected improvements of participants' knowledge or on their reactions in real situations. We can only guess about the positive impacts of the interactive algorithms from the fact that most of the student–authors did not have any difficulties launching their professional careers in acute medicine. Another improvement indicator can be inferred from the repeated successes of student–authors and student–players in international competitions of medical rescue teams.

Conclusions

The methodological aspects of our interactive algorithms for incorporation in the learning and teaching of acute medicine were presented. These interactive algorithms comprise the main part of the educational content of the AKUTNE.CZ portal and recently became the basis for a new extension for MEFANET, the education network of all medical faculties in the Czech Republic and Slovakia.

There are 25 algorithms in the Czech/Slovak and English languages, published online and covering a wide range of topics in acute medicine. The peer-reviewed algorithms were used for conducting PBL-like sessions in general medicine (first aid, anesthesiology and pain management, emergency medicine) as well as in nursing (emergency medicine for midwives, obstetric analgesia and anesthesia for midwives).

We investigated the students' perception of our interactive algorithms as an adjuvant to their medical and health care studies, especially in relation to clinical reasoning. The feedback from the survey among the AKUTNE.CZ users suggests that the students identify the interactive algorithms as an effective learning tool, serving to enhance their knowledge in the field of acute medicine. In addition, they expressed their keen desire to apply them not only in their leisure time, but also during face-to-face contact with their teachers at school or during clinical practice in the university hospital.

The AKUTNE.CZ interactive algorithms, as a software platform, are open to academic use worldwide. The already created and peer-reviewed algorithms, as simulation-based learning objects, can be included easily into any education website (subject to approval of the authors).

Conflicts of Interest

Daniel Schwarz is the principal investigator of the MEFANET clinical reasoning grant project, which funded development of the MEFANET e-publishing system as well as the interactive algorithms for teaching and learning acute medicine. The grant project also paid Daniel Schwarz, Petr Štourač, Martin Komenda, and Hana Harazim a small portion of their salaries at Masaryk University.

References

1. Holzinger A, Kickmeier-Rust MD, Wassertheurer S, Hessinger M. Learning performance with interactive simulations in medical education: Lessons learned from results of learning complex physiological models with the HAEMOdynamics SIMulator. Computers & Education 2009 Feb;52(2):292-301 [FREE Full text] [doi: 10.1016/j.compedu.2008.08.008]
2. Wong RW, Lochnan HA. A web-based simulation of a longitudinal clinic used in a 4-week ambulatory rotation: a cohort study. BMC Med Educ 2009;9:8 [FREE Full text] [doi: 10.1186/1472-6920-9-8] [Medline: 19187554]
3. Subramanian A, Timberlake M, Mittakanti H, Lara M, Brandt ML. Novel educational approach for medical students: improved retention rates using interactive medical software compared with traditional lecture-based format. J Surg Educ 2012;69(4):449-452. [doi: 10.1016/j.jsurg.2012.05.013] [Medline: 22677580]
4. Maloney S, Haas R, Keating JL, Molloy E, Jolly B, Sims J, et al. Breakeven, cost benefit, cost effectiveness, and willingness to pay for web-based versus face-to-face education delivery for health professionals. J Med Internet Res 2012;14(2):e47 [FREE Full text] [doi: 10.2196/jmir.2040] [Medline: 22469659]
5. Brett-Fleegler MB, Vinci RJ, Weiner DL, Harris SK, Shih MC, Kleinman ME. A simulator-based tool that assesses pediatric resident resuscitation competency. Pediatrics 2008 Mar;121(3):e597-e603 [FREE Full text] [doi: 10.1542/peds.2005-1259] [Medline: 18283069]
6. Pothiawala S, Lateef F. Hong Kong Journal of Emergency Medicine. 2012 Jan. Simulation training in emergency medicine (STEM): an integral component of residency curriculum URL: http://www.hkcem.com/html/publications/Journal/2012-1%20Jan/p41-45.pdf [accessed 2013-02-27] [WebCite Cache ID 6Ek9Md4ig]
7. Davids MR, Chikte UM, Halperin ML. Development and evaluation of a multimedia e-learning resource for electrolyte and acid-base disorders. Adv Physiol Educ 2011 Sep;35(3):295-306 [FREE Full text] [doi: 10.1152/advan.00127.2010] [Medline: 21908840]
8. Beux PL, Fieschi M. Virtual biomedical universities and e-learning. Int J Med Inform 2007;76(5-6):331-335. [doi: 10.1016/S1386-5056(07)00060-3] [Medline: 17407747]

XSL•FO

RenderX

9. Bamidis P, Kaldoudi E, Pattichis C. mEducator: A best practice network for repurposing and sharing medical educational multi-type content. In: Camarinha-Matos L, Paraskakis P, Afsarmanesh H, editors. Leveraging Knowledge for Innovation in Collaborative Networks. Berlin Heidelberg: Springer; 2009:769-776.

10. Kaldoudi E, Dovrolis N, Konstantinidis S, Bamidis P. Social networking for learning object repurposing in medical education. The Journal on Information Technology in Healthcare 2009;7(4):233-243 [FREE Full text]

11. Kaldoudi E, Dovrolis N, Konstantinidis ST, Bamidis PD. Depicting educational content repurposing context and inheritance. IEEE Trans Inf Technol Biomed 2011 Jan;15(1):164-170. [doi: 10.1109/TITB.2010.2092442] [Medline: 21134819]

12. Schwarz D, Dušek L. The MEFANET Project. URL: http://www.mefanet.cz/index-en.php [accessed 2013-02-27] [WebCite Cache ID 6EkA1pn4M]

13. Komenda M, Schwarz D, Feberová J, Stípek S, Mihál V, Dušek L. Medical faculties educational network: multidimensional quality assessment. Comput Methods Programs Biomed 2012 Dec;108(3):900-909. [doi: 10.1016/j.cmpb.2012.05.002] [Medline: 22640818]

14. Kleinpell R, Ely EW, Williams G, Liolios A, Ward N, Tisherman SA. Web-based resources for critical care education. Crit Care Med 2011 Mar;39(3):541-553. [doi: 10.1097/CCM.0b013e318206b5b5] [Medline: 21169819]

15. Creutzfeldt J, Hedman L, Heinrichs L, Youngblood P, Felländer-Tsai L. Cardiopulmonary resuscitation training in high school using avatars in virtual worlds: an international feasibility study. J Med Internet Res 2013;15(1):e9 [FREE Full text] [doi: 10.2196/jmir.1715] [Medline: 23318253]

16. Nathoo AN, Goldhoff P, Quattrochi JJ. Evaluation of an Interactive Case-based Online Network (ICON) in a problem based learning environment. Adv Health Sci Educ Theory Pract 2005 Aug;10(3):215-230. [doi: 10.1007/s10459-005-7851-3] [Medline: 16193402]

17. Servais EL, Lamorte WW, Agarwal S, Moschetti W, Mallipattu SK, Moulton SL. Teaching surgical decision-making: an interactive, web-based approach. J Surg Res 2006 Jul;134(1):102-106. [doi: 10.1016/j.jss.2005.11.583] [Medline: 16488433]

18. Roubidoux MA, Chapman CM, Piontek ME. Development and evaluation of an interactive Web-based breast imaging game for medical students. Acad Radiol 2002 Oct;9(10):1169-1178. [Medline: 12385511]

19. Dikshit A, Wu D, Wu C, Zhao W. An online interactive simulation system for medical imaging education. Comput Med Imaging Graph 2005 Sep;29(6):395-404. [doi: 10.1016/j.compmedimag.2005.02.001] [Medline: 15996851]

20. Holzinger A, Emberger W, Wassertheurer S, Neal L. Design, development and evaluation of online interactive simulation software for learning human genetics. Elektrotech. Inftech 2008 May;125(5):190-196. [doi: 10.1007/s00502-008-0537-9]

21. Štourač P, Křikava I, Štoudek R, Harazim H, Smékalová O, Kosinová M. AKUTNE.CZ educational portal. URL: http://www.akutne.cz/index-en.php [accessed 2013-02-27] [WebCite Cache ID 6EkA9tNl4]

22. MEFANET Central Gateway. URL: http://portal.mefanet.cz/index-en.php [accessed 2013-02-27] [WebCite Cache ID 6EkAEOHoF]

23. Czech academic identity federation eduID.cz. URL: http://www.eduid.cz/wiki/en/eduid/index [accessed 2013-02-27] [WebCite Cache ID 6EkAIeBd5]

24. MEFANET SANDBOX. URL: http://sandbox.mefanet.cz/index-en.php [accessed 2013-02-27] [WebCite Cache ID 6EkAM8VLQ]

25. WikiLectures. URL: http://www.wikilectures.eu/index.php/Main_Page [accessed 2013-02-27] [WebCite Cache ID 6EkAPMgtf]

26. Moodle-MEFANET e-learning courses. URL: https://moodle.mefanet.cz/ [accessed 2013-02-27] [WebCite Cache ID 6EkATjZqo]

27. MEFANET GAMES. URL: http://games.mefanet.cz/index-en.php [accessed 2013-02-27] [WebCite Cache ID 6EkAfPUZp]

28. SurveyMonkey. URL: http://www.surveymonkey.com/ [accessed 2013-02-27] [WebCite Cache ID 6EkAiwxIh]

29. Kofranek J, Matousek S, Rusz J, Stodulka P, Privitzer P, Matejak M, et al. The Atlas of Physiology and Pathophysiology: Web-based multimedia enabled interactive simulations. Comput Methods Programs Biomed 2011 Nov;104(2):143-153. [doi: 10.1016/j.cmpb.2010.12.007] [Medline: 21232813]

30. Thomas SR, Baconnier P, Fontecave J, Françoise JP, Guillaud F, Hannaert P, et al. SAPHIR: a physiome core model of body fluid homeostasis and blood pressure regulation. Philos Trans A Math Phys Eng Sci 2008 Sep 13;366(1878):3175-3197. [doi: 10.1098/rsta.2008.0079] [Medline: 18565814]

31. Iliescu R, Lohmeier TE. Lowering of blood pressure during chronic suppression of central sympathetic outflow: insight from computer simulations. Clin Exp Pharmacol Physiol 2010 Feb;37(2):e24-e33. [doi: 10.1111/j.1440-1681.2009.05291.x] [Medline: 19769610]

32. Cook DA, Triola MM. Virtual patients: a critical literature review and proposed next steps. Med Educ 2009 Apr;43(4):303-311. [doi: 10.1111/j.1365-2923.2008.03286.x] [Medline: 19335571]

Abbreviations

ALS: advanced life support
BLS: basic life support
CCM: critical care medicine

ICT: information and communication technology
ISSN: International Standard Serial Number
MEFANET: Medical Faculties Network
PBL: problem-based learning
SAML: Security Assertion Markup Language
SCCM: Society of Critical Care Medicine
XML: extensible markup language

XSL•FO
RenderX

Original Paper

An Easily Accessible Web-Based Minimization Random Allocation System for Clinical Trials

Lan Xiao[1], PhD; Qiwen Huang[1], MS; Veronica Yank[2], MD; Jun Ma[1,2], PhD, MD

[1]Research Institute, Palo Alto Medical Foundation, Palo Alto, CA, United States
[2]School of Medicine, Stanford University, Stanford, CA, United States

Corresponding Author:
Jun Ma, PhD, MD
Research Institute
Palo Alto Medical Foundation
Ames Building, 795 EL Camino Real
Palo Alto, CA
United States
Phone: 1 650 853 4809
Fax: 1 650 327 8309
Email: maj@pamfri.org

Abstract

Background: Minimization as an adaptive allocation technique has been recommended in the literature for use in randomized clinical trials. However, it remains uncommonly used due in part to a lack of easily accessible implementation tools.

Objective: To provide clinical trialists with a robust, flexible, and readily accessible tool for implementing covariate-adaptive biased-coin randomization.

Methods: We developed a Web-based random allocation system, MinimRan, that applies Pocock–Simon (for trials with 2 or more arms) and 2-way (currently limited to 2-arm trials) minimization methods for trials using only categorical prognostic factors or the symmetric Kullback–Leibler divergence minimization method for trials (currently limited to 2-arm trials) using continuous prognostic factors with or without categorical factors, in covariate-adaptive biased-coin randomization.

Results: In this paper, we describe the system's essential statistical and computer programming features and provide as an example the randomization results generated by it in a recently completed trial. The system can be used in single- and double-blind trials as well as single-center and multicenter trials.

Conclusions: We expect the system to facilitate the translation of the 3 validated random allocation methods into broad, efficient clinical research practice.

Introduction

Randomized controlled trials (RCTs) are the gold standard for assessing efficacy or effectiveness of biomedical and behavioral treatments. The ideal randomization procedure would achieve the following goals: (1) balanced arm sizes, (2) no selection bias (ie, unpredictability of future treatment assignment), and (3) no accidental bias (ie, low probability of confounding because of between-treatment imbalance in pretreatment characteristics of prognostic importance). However, no randomization procedure can achieve all these goals in every circumstance, which makes randomization conceptually straightforward but practically complex. Simple randomization—also called *unrestricted randomization*—minimizes selection bias but not accidental bias [1]. Hence, several restricted randomization procedures have been developed to address these limitations.

A practical solution that minimizes accidental bias when multiple prognostic factors are involved is the covariate-adaptive biased-coin randomization procedure widely known as the Pocock-Simon minimization method [2]. This method achieves marginal balance by accounting for all the selected pretreatment covariates for the previously assigned subjects and assigning the next subject to a treatment with a probability in favor of

minimizing the overall imbalance across the covariates. Use of nonextreme allocation probabilities (eg, 2/3:1/3 in a 2-arm trial) helps protect unpredictability [2,3]. A 2-way minimization method is another way to protect unpredictability by using probability to minimize either the "imbalance in the total numbers of subjects" or the "imbalance in the distribution of prognostic factors" [4]. Both Pocock-Simon and 2-way minimization methods only allow for balancing by categorical prognostic factors. However, categorizing continuous covariates may not always be feasible or preferable (eg, because of a lack of scientific basis for or consensus on cut points). Endo et al [5] extended the Pocock-Simon approach to incorporate continuous prognostic factors in 2-arm trials by using the symmetric Kullback-Leibler divergence (KLD) (ie, Jeffrey's divergence) index [6,7]. They demonstrated in a simulation study that, when continuous prognostic factors were included, the symmetric KLD method produced better covariate balance between treatments and more robust estimates of treatment effects than the Pocock-Simon method [5]. Despite their notable advantages and recommended use by many statistical and trialist commentators [8], these minimization methods remain infrequently used, to a large extent because of a lack of easily accessible tools [9].

In 2000, Kenjo et al [10] published their Web-based allocation system for multisite clinical trials using Pocock-Simon's minimization method, but as noted in Cai et al [11], that system did not appear to support multiple trials simultaneously or address blinding. Cai et al [11] subsequently developed a Web-based allocation system also based on the Pocock-Simon method specifically for double-blind trials (see subsequent definition). Although there is a freely available online directory of randomization software [12], only 2 downloadable programs of those listed, Minim and MinmPy [13], support minimization methods. QMinim, an online version of MinmPy, is also freely available [14]. However, none of these minimization tools include role management function. In other words, they do not allow for the granting of different access privileges to different users and, therefore, cannot support double-blind trials. In addition, each only offers a single minimization method.

To promote increased use of minimization methods in various study designs and settings, we have developed a robust Web-based randomization system, named MinimRan, with flexible and user-friendly features, including (1) choice of the minimization method (Pocock-Simon, symmetric KLD, or 2-way minimization), (2) differentiated access privileges for efficient user–project role management within and across projects within research teams, (3) simultaneous system access by multiple users within and across multiple sites, (4) convenient graphical user interfaces (GUI) for information input and output, (5) proper protections of blinding in single- and double-blind trials, (6) standardized reports for continuous, timely quality monitoring of the randomization process, and (7) interactive tools for information updates and error corrections.

Methods

System Design

We designed this Web-based random allocation system to support sequential covariate-balanced assignment of subjects in single-site and multisite trials that use single- or double-blind designs. Blinding helps prevent the subjects and/or researchers from biasing the outcome of a study. The definitions of single- and double-blind designs are described in the Multimedia Appendix 1.

As noted previously, our system's statistical algorithms are based on Pocock-Simon's minimization method (for trials with 2 or more arms), Endo et al's symmetric KLD minimization method [5] (currently limited to 2-arm trials), and 2-way minimization method (currently limited to 2-arm trials). All 3 methods can be applied to single-site or multisite studies. Users may create new projects and manage multiple existing projects, and may access comprehensive account management and monitoring functions—all within 1 account for the same research team.

Three-Tier System Architecture

The system uses a 3-tier architecture, which is the most widely used browser–server architecture. The 3-tier architecture consists of a presentation tier, logic tier, and data tier (Multimedia Appendix 1). The presentation tier is the user interface, which collects and displays information from the logic tier through a Web browser. The logic tier uses Tomcat server as the Web server and the Java application Java Server Page (JSP) along with Cascading Style Sheets (CSS) and JavaScript to build the Web application. The data tier is the back-end MySQL database server. Java Database Connectivity (JDBC) achieves database-independent connectivity between the Java programming language and the MySQL database. The detailed technical description of how the 3 tiers work together is included in Multimedia Appendix 1.

The system can be accessed by using Internet Explorer 8.0 or higher or Firefox [15].

User Roles

Three types of users—super, project manager, and general—can access the system with different types of privileges (Figure 1). Our technical team retains the role of the super user (and serves as the system administrator). The privileges of this role include (1) initiating study projects, (2) creating project manager accounts, (3) assigning projects to new or existing project managers, (4) supervising and ensuring proper uses of the system, and (5) planning for and responding to service outages and other system problems. After the super user authorizes an account for the project manager on a research team, the manager can then carry out the following project-specific activities: (1) defining project characteristics (eg, single- or double-blind trial, number of study groups, study sites, prognostic factors), (2) creating general user accounts with individual privileges specified, (3) deactivating general user accounts that are no longer needed, (4) performing randomization, (5) monitoring randomization with the ability to view and verify randomized records as appropriate to manager's blinding status (eg, only

XSL•FO
RenderX

masked numbers available for double-blind trial), (6) managing randomization results (eg, generating summary tables, downloading allocated records), and (7) updating project information (eg, adding study sites). General users on a given project can perform randomization and 1 or more of the other functions previously listed according to each person's privileges as assigned by the project manager. To help ensure blinding, the system will prompt the project manager to specify on a project-by-project basis which user(s) have permission to access the key that reveals subjects' group assignments. In a single-blind trial, the project manager and/or 1 or more general users may be granted permission. In a double-blind trial, the key should be accessed and kept only by a third party and not given to any researchers involved in the study, including the project manager and general users performing randomization, until the study is over. When the project manager designates a general user as a third party with permission to access the group assignment key (for details, see section Randomization Process and Blinding), the system will automatically disable all randomization-related functions for that person.

Creation and Maintenance of Research Projects

The steps for creating a new project are as follows (Figure 2):

1. Request to initiate a new project submitted by an existing or a new project manager. A brief description of the study must be provided that includes project name, purpose of study, beginning and expiration dates, funding source with grant number(s) if applicable, and applicant's contact information.

2. The super user will create a new project using the information provided and assign it to the project manager's account, which is also created at this point if there is not an existing account.

3. Definition of study parameters by the authorized project manager. The parameters include single- or double-blind trial, number of study groups and group names (optional), number and short names of study sites, projected maximum number of subjects for each study site (required for double-blind trials only), minimization method selected (Pocock-Simon, symmetric KLD, or 2-way minimization), biased assignment probability (not required for 2-way minimization), prognostic factors, and levels of each categorical factor. If the Pocock-Simon method is chosen, the user also needs to specify the number of initial subjects allocated using simple randomization (n=1 by default). The system recommends to users that they select simple randomization for the first 10 to 15 subjects as a strategy to prevent guessing of assignments when cases are few [16].

4. Creation of general user accounts and assignment of individual privileges by the project manager (for details, see section User Roles and Figure 1). General users are prompted to set their individual username and password when they log on for the first time.

Study projects can have 1 of 3 status designations: pending, ongoing, or expired. A project is pending when the authorized project manager has yet to complete steps 3 and 4 outlined previously. Once the setup is completed, randomization can then begin and the project's status changes to ongoing, and will remain as such until the expiration date specified during the project initiation process (see step 1). Thirty days before a project's expiration date, the system will generate an alert for the project manager who may at that point request an extension by emailing the super user. For expired projects, the project manager can view and download data records, but functions related to randomization of new subjects are deactivated. An expired project may be reactivated by the super user upon request.

Figure 1. User roles.
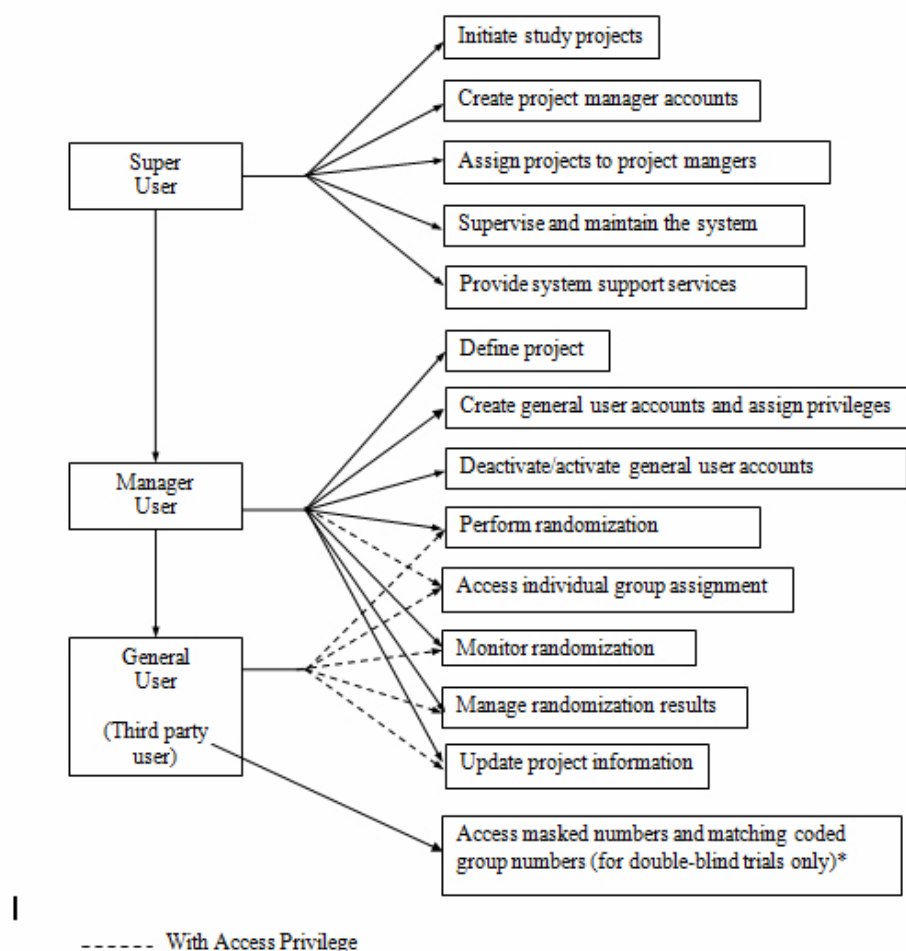
Figure 2. Steps to create a new project.
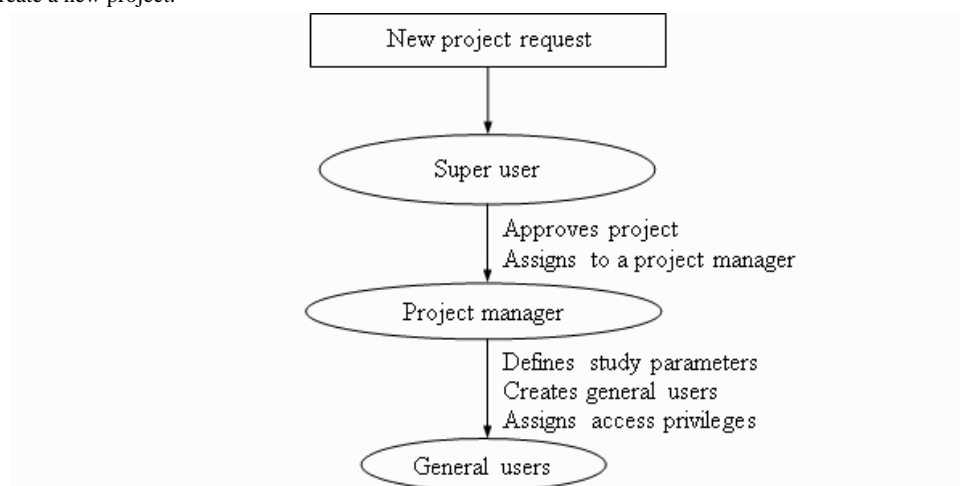
Randomization Process and Blinding

The system gives users the option of uploading records with subject IDs and prognostic factors for randomization by using a comma-separated values (CSV) data file or manually entering records 1 at a time. A CSV is a simple, widely supported file format in scientific, business, and consumer applications, and it permits efficient transfer of tabular data between programs.

Within a given trial, both input methods are available for the user to select during each randomization run, and switching between methods from 1 run to the next is permitted. With both methods, data validation before randomization is strongly encouraged in all cases. Specifically, the system prompts the user to verify the inputted subject information before executing the randomization. The system also automatically checks the values of the prognostic factors entered each time against

user-defined logic rules and generates an error message if any rule is violated. After data validation, the system opens the Pocock-Simon, symmetric KLD, or 2-way minimization method procedure depending on the user's selection. The system automatically generates random numbers and outputs the randomization results using system-generated coded group numbers (eg, 1, 2, or 3 for a 3-arm study) or group names (if the manager user opts to describe group numbers) for single-blind trials or using masked individual numbers (ie, system-generated random numbers with a preceding *M*) for double-blind trials. For the former, only users with permission to access the key that identifies the subjects (subject IDs provided by the research team) and to which group they belong (coded group numbers or group names) can see the randomization results. For the latter, the system generates a Masked_Num table upon completion of the project initiation steps (section Creation and Maintenance of Research Projects) and before randomization of the first subject in a double-blind trial. The table contains masked numbers and matching coded group numbers or group names (by study site if a multisite trial), which only a designated third-party general user can access (section User Roles) and download (as a CSV file) for encoding the treatments (eg, using masked numbers on drug bottle labels for distribution and tracking). The system provides project managers and general users performing randomization on double-blind trials with subjects' assigned masked numbers but not the associated group numbers. The user-projected maximum number of subjects to be enrolled plus 10% more determines the number of masked numbers generated by the system. The system will generate additional masked numbers if 90% of the initial set of numbers for any of the study group have been assigned. If the study includes multiple sites, this assignment will apply for each site. A designated user on a single-blind trial who is involved in conducting the research and has permission to access randomization results and the third party on a double-blind trial will be responsible for matching the randomization results and the actual study groups. As is standard practice in randomized clinical trials, this information should be kept in confidence (ie, not revealed to the researchers and participants who should remain blinded) until the study is ready to break the blind.

As previously mentioned, the system supports randomization at multiple sites and by multiple users. To prevent the race condition in a multi-user environment (ie, 2 or more users from the same study performing randomization tasks simultaneously), the system randomizes subjects in order of auto-incremented unique numbers that MySQL automatically generates when new records are inserted. The system also prevents duplicate randomization of the same subject ID within a project and will display an error message if this occurs. In spite of existing logical error checking provided by the system, some human entry errors may still be unavoidable. If the errors are found after randomization, the system only allows project managers to correct the entry errors and requires that he/she specify the reason, but the randomization results that happened before the corrections will remain unchanged. Randomization of any new subjects after the corrections, thus, will be based on the corrected information. The action of revision will be recorded and traceable in the randomization process data. Detailed

randomization process data (eg, study ID, factor values, random number, random probability) are captured in the back-end database and are retrievable to permit quality control and replication. A manager user with permission to access group assignments can download randomization process data for current and expired single-blind trials that he/she manages. For double-blind trials, however, the data can be requested from the super user only if the manager user attests in writing that a trial has broken the blind.

Back-End Database Design

The relational database built for MinimRan makes the system dynamic, flexible, scalable, and reliable The system uses MySQL to generate 8 tables for both single-blind and double-blind trials and 1 additional table for double-blinded trials only. The contents of each table and the relationships with other tables are described in detail in Multimedia Appendix 1.

Statistical Methods

Minimization is designed to minimize marginal imbalance over multiple important prognostic factors as each consecutive treatment assignment is made. The treatment assignment that results in the least overall imbalance will be chosen with a high probability (P_i), thereby increasing the chance of maximizing balance among the prognostic factors. The choice of P_i determines the degree of balance and the predictability of treatment assignment. Both Pocock-Simon and symmetric KLD methods define P_i as a fixed value throughout the whole or partial randomization stage, whereas 2-way minimization method defines dynamic P_i as a function of the imbalance in the total numbers of subjects. Depending on the type of prognostic factors chosen and user preference, in our system, users can choose one of these 3 methods for measuring imbalance.

Pocock-Simon's Imbalance Score

The first option for measuring imbalance is to use the Pocock-Simon minimization method, which requires that continuous prognostic factors be categorized to calculate treatment imbalance [2,17]. At an arbitrary point in the succession of randomizations and after the specified number of initial subjects for whom simple randomization is used is met, denote n_{ijm} as the number of patients with level m of factor j who have been previously assigned to treatment arm i (j=1,2,...,J; m_j= 1,2,...,M_j; and i=1,2,..., I, where J, M_j, and I are the numbers of prognostic factors, levels of factor j, and treatment arms, respectively). Let the next participant entering the trial have levels r_1, r_2,...,r_J on the prognostic factors 1,...,J. Pocock and Simon proposed several ways of measuring the cumulative imbalance on the previously assigned subjects and after assignment of a new participant [2,17]. We chose to balance the marginal treatment totals for each level of each patient factor in our system [17]. Figure 3 displays the equation used, where G_i is the marginal treatment total if the new participant is assigned to treatment i. The G scores corresponding to each treatment i are then ranked from the smallest to the largest and assigned with the corresponding P_i.

The Symmetric KLD Index

The second option for measuring imbalance is to use the algorithm that measures the amount of imbalance between treatments (currently limited to 2) over multiple prognostic factors by computing a symmetric KLD index after a permuted block of the first 4 subjects have been assigned [5]. Let treatment be coded i (i=1,2). Consider any arbitrary point with the number of subjects n>4. Let x_{ijk} be the value of kth (k=1,2,...,n_i) participant assigned to treatment i with the jth (j=1,2,...,j') continuous prognostic factor, and p_{ijm} be the proportion of subjects assigned to the level m (m=1,2,...,M) of the jth (j+1,...,J) categorical prognostic factor. The difference in the distribution of prognostic factors between 2 treatments i and i' (d_i) can be measured as shown in the equation in Figure 4.

When the new participant n+1 is enrolled, d_i is calculated by assuming that this individual is allocated to i where i can be either treatment. Hence, the total number of subjects for treatment i becomes n_i +1 and the number of subjects for the other treatment i' (n_i') remains unchanged. The value d_i represents the amount of imbalance in treatment i assuming the new subject is allocated to this treatment. The higher probability P_i is then assigned to the treatment arm with lower d_i. The symmetric KLD algorithm assumes a multivariate normal distribution for continuous prognostic factors, although Endo et al [5] demonstrated that the algorithm was robust to nonnormally distributed data. If the symmetric KLD method is chosen, the system displays a message to alert the user to the multivariate normal distribution assumption and advises consulting a biostatistician on the need for data transformation if it is believed that serious violations may occur given prior knowledge of the expected distributions of the continuous factors used in the trial.

Two-Way Minimization Method

This method (currently limited to trials with 2 arms) calculates the imbalance in the total numbers of subjects and the imbalance in the distributions of prognostic factors. It then chooses, based on the defined probability P_i, to minimize either 1 of these 2 imbalances.

Consider an arbitrary point in the trial after a simple randomization scheme allocates at least 1 subject in each group. Let n_T and n_C denote the total numbers of subjects allocated to the treatment group and the control group.

For the equation used for imbalance in the total numbers of subjects, see Figure 5. For the equation used for imbalance in the distributions of prognostic factors, see Figure 6.

We define probability P to determine that the new subject is allocated to minimize delta with probability=P and to minimize D with probability=$1-P$, where P is chosen based on the original paper-proposed function: $P=1-0.95^{\delta}$.

Figure 3. Equation for marginal treatment total.

$$G_i = \sum_{j=1}^{J} \sum_{m=r_j} n_{ijm}$$

Figure 4. Equation for the KLD index.

$$d_i = 1/2 \sum_{j=1}^{j'} \{(\overline{x_{ij}} - \overline{x_{i'j}})^2 + (V_{ij} + V_{i'j})\}\left(\frac{1}{V_{ij}} + \frac{1}{V_{i'j}}\right) - 2j' + \sum_{j=j'+1}^{J} \sum_{m=1}^{M_j} (p_{ijm} - p_{i'jm}) \log \frac{p_{ijm}}{p_{i'jm}}$$

$$\text{Where } \overline{x_{ij}} = \frac{1}{n_i} \sum_{k=1}^{n_i} x_{ijk} \text{ and } V_{ij} = \frac{1}{n_i-1} \sum_{k=1}^{n_i} (x_{ijk} - \overline{x_{ij}})^2.$$

Figure 5. Equation for imbalance in the total numbers of subjects for two-way minimization method.

$$\delta = |n_T - n_C|$$

Figure 6. Equation for imbalance in the distributions of prognostic factors for two-way minimization method.

$$D = \sum_{j=1}^{J} \left(\sum_{m=1}^{M_j} (|p_{Tjm} - p_{Cjm}|) \right)/M_j$$

Results

All the functions in the system that we describe here have been fully tested in 2 popular Web browsers (ie, Internet Explorer 8.0 and Firefox) and already implemented in actual RCTs, 1 of which is a recently completed 3-arm study, Evaluation of Lifestyle Interventions to Treat Elevated Cardiometabolic Risk in Primary Care (E-LITE; ClinicalTrials.gov NCT00842426). E-LITE was designed to evaluate 2 behavioral weight-management interventions compared with usual care, in 1 primary care clinic of a large multispecialty group practice

XSL•FO
RenderX

in Northern California [18]. The protocol specifies 7 prognostic factors for randomization: age, gender, race, pretrial online access to personal health records, fasting blood glucose, body mass index, and waist circumference. The Pocock-Simon minimization method was used. The summary table of all randomized records (n=241), which was generated by the Web-based system (with the exception of the *P* values), shows better than chance balance across all 7 prognostic factors among the 3 treatment arms (Table 1).

Table 1. Between-group differences in prognostic factors for the Evaluation of Lifestyle Interventions to Treat Elevated Cardiometabolic Risk in Primary Care (E-LITE) study.

Factor and level	Number in each treatment			Total	Max group difference	P value
	1	2	3			
Age (years)						
18-44	21	19	20	60	2	.84
45-64	52	46	48	146	6	
65-79	8	13	12	33	5	
80-100	0	1	1	2	1	
Blood glucose (mg/dL)						
0-99	39	34	37	110	5	.41
100-109	31	30	31	92	1	
110-119	7	11	13	31	6	
120-125	4	4	0	8	4	
Body mass index (kg/m^2)						
25-29.9	38	37	38	113	1	.99
30-34.9	25	27	24	76	3	
35-39.9	12	9	11	32	3	
≥40	6	6	8	20	2	
Gender						
Male	44	41	44	129	3	.94
Female	37	38	37	112	1	
Access to personal health records						
No	15	15	18	48	3	.81
Yes	66	64	63	193	3	
Race						
Hispanic	2	4	4	10	2	.93
Asian/Pacific Islander	14	13	14	41	1	
Non-Hispanic white	64	61	63	188	3	
Other	1	1	0	2	1	
Waist circumference (cm)						
<37	9	10	11	30	2	.99
37-<40	23	19	21	63	4	
40-<42	15	16	14	45	2	
≥42	34	34	35	103	1	
Total	81	79	81	241	2	

XSL•FO
RenderX

Discussion

Principal Findings

We have developed a Web-based randomization system to facilitate use of the Pocock-Simon, symmetric KLD, and 2-way minimization methods. It provides user-friendly and error-resistant Web interfaces that are applicable to single- and double-blind trials as well as single-center and multicenter trials.

Randomization ensures that research subjects are assigned to a treatment independent of baseline characteristics, measured or unmeasured, including characteristics that are the current values of potential outcomes of interest. Minimization as an adaptive randomization procedure has the desirable features of minimizing accidental bias while maximizing the precision of treatment effect estimates, particularly in small trials [8,19]. Given that methods to improve the prospects for balance increase the risk of selection bias [20] and the nature of the trade-off depends on the details (eg, masking or not, knowledge or ignorance of baseline prognostic factors), the proper choice of biased assignment probability P_i specifically for Pocock-Simon and symmetric KLD methods varies according to individual study circumstances [2,21]. Our Web-based randomization system incorporates Efron's biased-coin principle [3] and allows users to specify P_i when defining a new project and adjust it after project initiation if warranted (eg, if the initial P_i leads to imbalance measures exceeding a prespecified threshold in a given study).

Interactions between prognostic factors may affect response to treatment. It would be impractical to balance for all covariate interactions of any order in most clinical trials [2]. Nevertheless, all 3 minimization methods included in MinimRan can incorporate a first-order interaction between 2 categorical prognostic factors by creating a new variable whose levels correspond to all combinations of the 2 factors [2]. For example, a variable indicating gender (female, male) by smoking status (smoker, nonsmoker) interaction would have 4 levels: (1) female and smoker, (2) female and nonsmoker, (3) male and smoker, and (4) male and nonsmoker. Additionally, Endo et al's symmetric KLD method [5] can also account for first-order interactions between a categorical and a continuous prognostic factor and between 2 continuous prognostic factors.

Decisions regarding appropriate P_i values specifically for Pocock-Simon and symmetric KLD methods and interaction terms between prognostic factors need to be study specific and should only be made by experienced researchers, preferably in consultation with a qualified biostatistician. The final selections should be clearly documented in the study protocol.

For a 2-arm study, the system provides the option of using the Pocock-Simon method [2], Endo et al's symmetric KLD method [5], or the 2-way minimization method [4]. Although the KLD method has the advantage of permitting continuous and categorical prognostic factors and the 2-way minimization method protects unpredictability of new subject allocation, both algorithms are currently limited to randomization in 2-arm trials [4,5]. In contrast, the Pocock-Simon method can accommodate RCTs with more than 2 arms.

To support potential users of the MinimRan Web-based randomization system, we provide an online Q&A page and downloadable user manual, as well as a test-run option using dummy data (up to 10 subjects). In addition, we provide users with the option of contacting our development and super user team regarding tailoring characteristics of the program to their specific needs. For example, although the flexibility of our stand-alone system allows it to be used at institutions that do not yet have electronic data capture (EDC) or clinical trial management (CTM) systems, there may be users who want our randomization system to be integrated within their EDC and/or CTM systems, which are becoming more frequent in industry trial settings [22]. Similarly, academic and private or public health care centers that less frequently have EDC/CTM systems, but that use electronic health record (EHR) systems and also conduct clinical trials may seek our assistance to connect our Web-based randomization system to their EHR. There are administrative and regulatory requirements, however, that make such integration with the EHR challenging but not impossible [23]. For instance, integration requires institutional review board (IRB) authorization and strict adherence to the Health Insurance Portability and Accountability Act (HIPAA) privacy and security rules, which entail considerable effort. Furthermore, because many different EDC, CTM, and EHR systems exist, a Web-based stand-alone randomization system, such as the one described here, that is readily adaptable to different potential contexts of use has important practical value.

Conclusion

The Web-based randomization system (MinimRan) described in this paper provides clinical trialists with a robust, flexible, and readily accessible tool for implementing covariate-adaptive biased-coin randomization. We have presented the system's essential statistical and computer programming features and provided an example of the randomization results that it generated in 1 of our recent RCTs. A tool such as this can facilitate translation of validated randomization methods into broad, efficient use in clinical research.

XSL·FO

RenderX

interpretation of the data; or preparation, review, or approval of the manuscript. The authors would like to acknowledge Sarah Knowles in the Palo Alto Medical Foundation Research Institute for testing the system and providing helpful comments and suggestions for online Q&A and user manuals.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Technical details for the Web system and instruction for randomizing multi-site double-blind trial.

[PDF File (Adobe PDF File), 157KB - jmir_v15i7e139_app1.pdf]

References

1. Carter BR, Hood K. Balance algorithm for cluster randomized trials. BMC Med Res Methodol 2008;8:65. [doi: 10.1186/1471-2288-8-65]
2. Pocock SJ, Simon R. Sequential treatment assignment with balancing for prognostic factors in the controlled clinical trial. Biometrics 1975 Mar;31(1):103-115. [Medline: 1100130]
3. EFRON B. Forcing a sequential experiment to be balanced. Biometrika 1971;58(3):403-417. [doi: 10.1093/biomet/58.3.403]
4. Chen LH, Lee WC. Two-way minimization: a novel treatment allocation method for small trials. PLoS One 2011;6(12):e28604 [FREE Full text] [doi: 10.1371/journal.pone.0028604] [Medline: 22163317]
5. Endo A, Nagatani F, Hamada C, Yoshimura I. Minimization method for balancing continuous prognostic variables between treatment and control groups using Kullback-Leibler divergence. Contemp Clin Trials 2006 Oct;27(5):420-431. [doi: 10.1016/j.cct.2006.05.002] [Medline: 16807130]
6. Kullback S. Information theory and statistics. In: Information Theory and Statistics. Mineola, NY: Dover Publications; 1997.
7. Rubner Y, Tomasi C, Guibas LJ. The Earth Mover's Distance as a metric for image retrieval. International Journal of Computer Vision 2000;40(2):99-121 [FREE Full text] [doi: 10.1023/A:1026543900054]
8. Scott NW, McPherson GC, Ramsay CR, Campbell MK. The method of minimization for allocation to clinical trials. a review. Control Clin Trials 2002 Dec;23(6):662-674. [Medline: 12505244]
9. Cai HW, Xia JL, Gao DH, Cao XM. Implementation and experience of a web-based allocation system with Pocock and Simon's minimization methods. Contemp Clin Trials 2010 Nov;31(6):510-513. [doi: 10.1016/j.cct.2010.07.009] [Medline: 20674777]
10. Kenjo Y, Antoku Y, Akazawa K, Hanada E, Kinukawa N, Nose Y. An easily customized, random allocation system using the minimization method for multi-institutional clinical trials. Comput Methods Programs Biomed 2000 May;62(1):45-49. [Medline: 10699684]
11. Cai H, Xia J, Xu D, Gao D, Yan Y. A generic minimization random allocation and blinding system on web. J Biomed Inform 2006 Dec;39(6):706-719. [doi: 10.1016/j.jbi.2006.02.015] [Medline: 16616696]
12. Directory of randomisation software and services. URL: http://www-users.york.ac.uk/~mb55/guide/randsery.htm [accessed 2013-04-09] [WebCite Cache ID 6FlFHV5Cn]
13. Saghaei M, Saghaei S. Implementation of an open-source customizable minimization program for allocation of patients to parallel groups in clinical trials. Journal of Biomedical Science and Engineering 2011;4:734-739. [doi: 10.4236/jbise.2011.411090]
14. QMinim. Online minimization URL: http://qminim.saghaei.net/ [accessed 2013-04-09] [WebCite Cache ID 6FlGnhHd9]
15. Xiao L, Huang Q, Yank V, Ma J. Palo Alto Medical Foundation Research Institute. DHSR/PAMFRI MinimRan (Beta) URL: http://studies.pamfri.org:8080/minimRan/index.jsp [accessed 2013-07-01] [WebCite Cache ID 6HmziGgGX]
16. Zielhuis GA, Straatman H, van 't Hof-Grootenboer AE, van Lier HJ, Rach GH, van den Broek P. The choice of a balanced allocation method for a clinical trial in otitis media with effusion. Stat Med 1990 Mar;9(3):237-246. [Medline: 2343219]
17. Pocock SJ. Clinical Trials: A Practical Approach. Chichester, UK: John Wiley and Sons; 1991.
18. Ma J, Yank V, Xiao L, Lavori PW, Wilson SR, Rosas LG, et al. Translating the Diabetes Prevention Program lifestyle intervention for weight loss into primary care: a randomized trial. JAMA Intern Med 2013 Jan 28;173(2):113-121. [doi: 10.1001/2013.jamainternmed.987] [Medline: 23229846]
19. McEntegart DJ. The Pursuit of Balance Using Stratified and Dynamic Randomization Techniques: An Overview. Drug Information Journal 2003 Jul 01;37(3):293-308. [doi: 10.1177/009286150303700305]
20. Taves DR. Minimization does not by its nature preclude allocation concealment and invite selection bias, as Berger claims. Contemp Clin Trials 2011 May;32(3):323. [doi: 10.1016/j.cct.2010.12.010] [Medline: 21187164]
21. Hagino A, Hamada C, Yoshimura I, Ohashi Y, Sakamoto J, Nakazato H. Statistical comparison of random allocation methods in cancer clinical trials. Control Clin Trials 2004 Dec;25(6):572-584. [doi: 10.1016/j.cct.2004.08.004] [Medline: 15588744]

XSL•FO
RenderX

22.   Harper B. Bio-IT world. 2007 Feb 12. Meshing EDC with CTMS URL: http://www.bio-itworld.com/issues/2007/feb/cover-story/ [accessed 2013-06-20] [WebCite Cache ID 6HWebIZ36]
23.   Fiore LD, Brophy M, Ferguson RE, D'Avolio L, Hermos JA, Lew RA, et al. A point-of-care clinical trial comparing insulin administered using a sliding scale versus a weight-based regimen. Clin Trials 2011 Apr;8(2):183-195 [FREE Full text] [doi: 10.1177/1740774511398368] [Medline: 21478329]

Abbreviations

CSS: Cascading Style Sheets
CSV: comma-separated values
CTM: clinical trial management
EDC: electronic data capture
EHR: electronic health record
GUI: graphical user interfaces
JDBC: Java Database Connectivity
JSP: Java Server Page
KLD: Kullback–Leibler divergence
RCT: randomized controlled trial

XSL•FO
RenderX

XSL•FO
RenderX